RESEARCH PROTOCOL

Development of STARD for Abstracts:

Essential items in reporting diagnostic accuracy studies in journal or conference abstracts

Version: April 23, 2015

Jérémie Cohen, Daniël Korevaar, Lotty Hooft, Johannes B. Reitsma and Patrick Bossuyt

Contact j.f.cohen@amc.uva.nl

EXECUTIVE COMMITTEE

Jérémie F. Cohen

Daniël A. Korevaar

Patrick M. Bossuyt

Academic Medical Center – University of Amsterdam

Department of Clinical Epidemiology, Biostatistics and Bioinformatics

Meibergdreef 9, P.O. Box 22660

1100 DD Amsterdam, The Netherlands

Phone: 0031-20-5668975 Fax: 0031-20-6912683

Lotty Hooft

Johannes B. Reitsma

Dutch Cochrane Centre

University Medical Center Utrecht – University of Utrecht

Julius Center for Health Sciences and Primary Care

Huispost Str. 6.131, P.O. Box 85500

3508 GA Utrecht, The Netherlands

Phone: 0031-88-7568003 Fax: 0031-88-7568099

ABSTRACT

Diagnostic accuracy studies evaluate the ability of a diagnostic test to correctly identify patients as having or not having a particular disease. Readers of diagnostic accuracy study reports use abstracts to decide whether they should look for the full study report and invest time in reading it. This decision requires an informative description of the purpose, methods and results of the study. However, abstracts of diagnostic accuracy studies often insufficiently report such information, making it difficult for readers to assess the validity of the study findings. In this protocol, we explain the rationale for a new reporting guideline for abstracts of diagnostic accuracy studies and describe how we plan to develop it.

TABLE OF CONTENTS

GLOSSARY

Executive Committee

This consists of the principal investigators (DAK and JFC, respectively doctoral and post-doctoral research fellows), and the supervising team (PMB, JBR and LH, respectively Professor in Clinical Epidemiology, Associate Professor in Clinical Epidemiology, and co-director of the Dutch Cochrane Centre).

Advisory Board

Experts with knowledge in the field of diagnostic accuracy studies and reporting guidelines. These consist of current members of the STARD Steering Committee.

Research team

The combination of the Executive Committee and Advisory Board.

Delphi participants

Participants in the Delphi survey. These consist of current members of the STARD group.

Checklist

A one-page document listing the essential items that should be reported in all abstracts of diagnostic accuracy studies. It constitutes the core of the reporting guideline.

Statement

Provides the rationale for the initiative of developing this reporting guideline.

Explanation and Elaboration (E&E)

Provides explanations and examples of reporting for each item in the checklist.

Reporting guideline

The combination of the checklist, statement and E&E material.

BACKGROUND AND OBJECTIVES

1. Diagnostic accuracy studies

Diagnostic accuracy studies evaluate the performance of one or more medical tests in correctly classifying study participants as having a target condition or not. The aim of classifying patients can be diagnosis, staging, prognosis, or prediction. The tests whose accuracy is evaluated are referred to as index tests. Evaluating the performance of a test in correctly classifying patients is typically done by comparing the index test results with those of the clinical reference standard (sometimes referred to as the "gold" standard). This is the best available clinical method for detecting the target condition. The reference standard can be a single test, procedure, or observation, a combination of these, or a judgment by a group of experts.

Diagnostic accuracy studies report the distribution of the index test results against the results obtained with the reference standard, usually in a cross tabulation. This can be used to estimate the index test's sensitivity (the proportion of participants with the target condition correctly identified as such by the index test) and specificity (the proportion without the target condition correctly identified as such by the index test). Several other accuracy statistics are also used, such as the negative and positive predictive value of the test, or area under the receiver operating characteristic curve.

2. Elements of validity in diagnostic accuracy studies

Internal validity refers to whether the results of the study are prone to bias, i.e. if the results of the results of the study do not reflect the "true" accuracy of the test. Major sources of bias in

diagnostic accuracy studies include methodological flaws in participant recruitment, data collection, test execution and interpretation, and data analysis [1].

External validity refers to whether the results of a study, although possibly unbiased, are not applicable to another setting and population of interest. Diagnostic accuracy varies across studies because of variations in study setting, participant characteristics, disease prevalence and severity, and aspects of test execution and interpretation [1].

3. Standards for the Reporting of Diagnostic Accuracy Studies (STARD)

In 2003, the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) statement was first published [2]. STARD provides guidance for the reporting of diagnostic accuracy studies. The STARD statement contains a checklist of 25 items that should be presented in all reports of diagnostic accuracy studies, covering key elements from study design and setting, selection of participants, execution and interpretation of the index test and reference standard, data analysis, and presentation of results. The guiding principle in the development of the STARD checklist was to select items that would help readers to appraise elements of internal and external validity of diagnostic accuracy studies. STARD has significantly improved the completeness of reporting of diagnostic accuracy studies, although it remains suboptimal [3]. The STARD checklist is currently being updated by the STARD group. The updated checklist should be released in 2015.

4. Incomplete reporting of journal abstracts of diagnostic accuracy studies

Unlike some other reporting guidelines, such as CONSORT (for randomized controlled trials) [4] and PRISMA (for systematic reviews) [5], STARD so far has not provided guidance for writing

abstracts. In a recent literature survey of the diagnostic accuracy studies published in twelve high-impact journals in 2012, we evaluated the completeness of reporting of 103 abstracts using 21 items deemed essential based on published guidance for adequate reporting and study validity assessment [6]. These items focused on study identification, rationale, objectives, methods for recruitment and testing, participant baseline characteristics, missing data, test results and reproducibility, estimates of diagnostic accuracy, and discussion of study findings, implications and limitations. Many abstracts were found to be insufficiently informative. Essential information on study participants, setting, participant sampling, blinding, and confidence intervals around accuracy estimates were only reported in less than half of the abstracts. The mean number of reported items per abstract was 10.1 (SD 2.2; range 6 to 15).

During the STARD update meeting held in Amsterdam in September 2014, the STARD Steering Committee agreed that the development of an official extension of STARD to address the issues of reporting in abstracts of diagnostic accuracy studies has high priority.

5. Mission statement

Who we are?

This project is being developed on behalf of the STARD Group, an international collaboration of methodologists, researchers involved in the conduct, design and analysis of studies or systematic reviews of diagnostic accuracy, and journal editors. The Executive Committee for this project consists of 2 junior researchers (JFC and DAK) and 3 senior researchers (LH, PMB and JBR) with expertise in diagnostic research methodology. PMB and JBR were part of the team who developed the original STARD statement; all members are part of the STARD updating process.

What we want to achieve?

Our goal is to develop, disseminate and implement a robust reporting guideline that will help improve the informativeness of abstracts of diagnostic accuracy studies regarding essential items that are needed to evaluate elements of study validity. Readers use abstracts to make decisions about whether they should look for the full study report and invest time in reading the full paper. When screening abstracts, readers should be able to appraise elements of validity of the study in terms of bias and applicability. This can only be achieved if abstracts are sufficiently informative. Our long term goal is to increase the usability of available diagnostic research evidence.

How will the guideline be helpful to our target audience?

Our primary target audience consists of authors of diagnostic accuracy study reports. We aim to help them in the writing phase of the abstract of the report. The new guideline will help authors by offering them a checklist containing essential items that should be in every diagnostic accuracy study abstract. We will also provide material to explain the scientific rationale for each recommended item.

6. Specific objectives

With respect to our mission statement, our specific objectives are:

- to identify a set of potentially relevant items for the reporting of diagnostic accuracy studies in abstracts (stage 1);
- to develop a consensual list of essential items that should be reported in all abstracts of diagnostic accuracy studies (stage 2);
- to develop a checklist, a statement and E&E online material based on the list (stage 3);
- to develop a specific plan for activities that should be undertaken to improve informative reporting of abstracts, using the list (stage 4).

METHODS

This protocol has been inspired by the "Guidance for Developers of Health Research Reporting Guidelines" [7], the methods used for the development of CONSORT for Abstracts [4], and publications related to the development of the "Guideline for Reporting Evidence based practice Educational interventions and Teaching" (GREET) [8, 9].

1. Definition of essential item

Our goal is to develop a list of essential items that should be reported in all abstracts of diagnostic accuracy studies. The guiding principle in doing so is to select essential items to help readers to appraise the potential for bias and the applicability of the study findings. The other consideration that will shape the selection of items is the usual word limit of 200-300 words in journal and conference abstracts. Therefore, we aim to develop a checklist with a maximum of about 15 items, as CONSORT for Abstracts [10].

2. Stage 1: Item generation and pre-selection

As part of an evaluation of the reporting quality of abstracts of diagnostic accuracy studies (see Background and Objectives, paragraph 4), 4 authors (DAK, JFC, LH and PMB) generated a list of 36 potentially essential items. This list of items was based on the STARD statement [2], the CONSORT for Abstracts checklist [4], the PRISMA for Abstracts checklist [5], QUADAS-2 [11], existing guidance on the structured reporting and the assessment of the quality of journal abstracts in general [12-14], and previous studies evaluating the content of abstracts of diagnostic accuracy studies [15, 16]. This list was updated using the draft checklist that was prepared following the consensus STARD update meeting held in Amsterdam in September 2014, for a

total of 39 items. From this list of 39 items, the Executive Committee agreed to remove 12 items that seemed not essential for STARD for Abstracts (for example details about sample size calculations). We ended up with a list of 27 items to submit to Delphi participants.

3. Stage 2: Delphi survey

3.1. Design

We will conduct a modified Delphi survey according to the methods described by Philips *et al.* for the GREET reporting guideline [8], with slight adaptations. The Delphi method is "a structured process of obtaining opinion from a group of experts by means of a series of questionnaires, each one refined based on the feedback from respondents on a previous version" [7]. In line with the development of CONSORT for Abstracts [4], the Delphi survey will comprise a series of three rounds of questionnaires and feedback to the group [17]. The aim of the Delphi survey is to obtain consensus on essential items that should be on the list.

3.2. Delphi facilitators

Two authors (JFC and DAK) will be the Delphi facilitators. For each round of the survey, they will prepare the questionnaires, send invitations, monitor responses, send reminders, collect, analyze and process responses, and prepare feedback for the next round.

3.3. Delphi participants

All members of the STARD Group (85 people) will be invited to participate in the Delphi survey. The STARD Group brings together clinical epidemiologists and statisticians involved in the

development of methods for studies of diagnostic accuracy, researchers directly involved in the conduct, design and analysis of studies or systematic reviews of diagnostic accuracy, and editors of journals that frequently publish studies of diagnostic accuracy or with experience in developing reporting guidelines.

Invitation and participation in the Delphi survey will be completed via email. Invitees will have 3 weeks to respond to the initial invitation to participate in the process. There will be one reminder for the invitation to participate two weeks after the initial one. All of those who accept to participate in the Delphi survey will be invited to complete each Delphi round, regardless of participation in the previous round, unless they indicate their will to be withdrawn from the Delphi process.

3.4. Online surveys

The Delphi survey will be conducted using SurveyMonkey©, an electronic online survey facility. For each round, Delphi participants will have 3 weeks to respond, and one reminder will be sent out two weeks after the initial invitation.

3.5. Rating

At each round of the Delphi survey, participants will be instructed to rate to what extent each item is essential, on a 5-point Likert scale, a rating of 1 meaning that the item is absolutely not required in all abstracts, and a rating of 5 meaning that the item is absolutely essential for reporting and cannot be absent from any abstract. Likert scores will then be designated into three categories [4]:

- 1–2: Low score: item should not be included in the checklist;
- 3: Moderate score: item should be discussed for inclusion in the checklist;
- 4–5: High score: item should be included in the checklist.

3.6. Consensus criterion

The consensus criterion will be defined based on agreement on rating. For an item to be considered to have met the consensus criterion, ≥2/3 (66.7%) of Delphi participants will need to rate an item within one of the above mentioned 3 categories (low, moderate or high score).

3.7. First round of the survey

The first round will focus on the items generated in Stage 1. Items will be ordered under headings that correspond to the conventional sections of a structured biomedical research abstract: title, background and objectives, methods, results, and discussion and conclusions. Delphi participants will be invited to rate each item using the 5-point Likert scale, with the option to provide a brief justification or reference to support their rating. They will also have the opportunity to suggest additional items. The survey will end with an open comments box.

3.8. Second round of the survey

Delphi participants will be provided with descriptive feedback for each item (mean score with SD and range, percent agreement and consensus [yes/no]). Items that reach consensus in the first round will not be required for further comment in the second round of the survey. All other items will be categorized and re-ordered based upon their mean score in the first round (high 4–5, moderate 3, low score 1–2). New items suggested by Delphi participants in the first round will

also be listed. Again, participants will be asked to rate each item using the 5-point Likert scale. The statistical analysis will repeat that from the first round.

3.9. Third round of the survey

In this round the process from the previous two rounds of the survey will be repeated. Items that reached consensus in the first or second round will not be required for further comment. Items repeatedly scored as not being essential (i.e. those with a mean score of 1–2) in the first 2 rounds of the survey will be removed from the checklist and will not be required for rating in the third round.

3.10. Output of the Delphi survey

At the completion of the third round, all items will have been described as follows:

	Mean score (SD)			Distribution of scores in the last round			
	Round 1	**Round 2**	**Round 3**	**1–2**	**3**	**4–5**	**Consensus**
Item 1	X (SD)	X (SD)	X (SD)	A%	B%	C%	(Yes/No)
Item 2							
…							

The following course of action will then be considered for each item:

	Consensus	
Mean rating	**No**	**Yes**
Low score (1–2)	Discuss inclusion	Do not include in the checklist
Moderate score (3)	Discuss inclusion	Discuss inclusion
High score (4–5)	Discuss inclusion	Include in the checklist

This will constitute the basis for the content of the first draft of the STARD for Abstracts checklist.

4. Stage 3: Development of the checklist, statement and E&E material

Following the Delphi survey, the Executive Committee will draft the initial STARD for Abstracts checklist and statement. The draft documents will be provided to the Advisory Board who will be convened to participate to an online forum to discuss its content, and later on, in a live consensus discussion on Skype©. The live consensus discussion will provide the opportunity for the Research Team to discuss the content of the abstract checklist and statement, directions for the development of specific E&E material, and plans for dissemination and implementation. At this stage, members of the Research Team will still be able to suggest additions, subtractions, and edits to the checklist.

The STARD for Abstracts checklist and statement will then undergo iterative drafting via email circulation into the Executive Committee until consensus. In parallel, the Executive Committee will prepare specific E&E material that should later be published online (i.e. explanations of what is meant by each item, short rationale for reporting each item, and examples of complete reporting in abstracts).

5. Stage 4: Publication and post-publication activities

A strategy for dissemination and implementation of STARD for Abstracts will be discussed and agreed upon during the consensus discussion. This will include publication of the checklist and statement, preparation and online publication of the E&E material, solicitation of journal editors to actively and explicitly endorse the new reporting guideline, and identification of opportunities to increase awareness among editors, peer reviewers and researchers by publishing editorials. We will also encourage translations of the guideline.

The checklist, the statement and E&E material will be submitted for approval to the Advisory Board before being submitted for publication or put online.

OTHER INFORMATION

1. Registration

The final protocol of this study will be registered on the EQUATOR website (http://www.equator-network.org).

2. Ethical considerations

In the Netherlands, non-interventional questionnaires do not fall under the scope of the Medical Research Involving Human Subjects Act and thus do not require approval by a medical ethics committee (http://www.ccmo.nl/en/questionnaire-research).

3. Funding

No specific funding is needed for this project.

4. Contributions

JFC wrote the first draft of the protocol. DAK edited the protocol several times. PMB, HR and LH provided supervision and critical comments. We also thank Prof. Constantine A. Gatsonis, PhD (Department of Biostatistics, Brown University School of Public Health, Providence, USA) for his valuable feedback and advise on this protocol.

5. Provisional timeline

Item generation (stage 1)	*Completed*
Delphi survey (stage 2)	April – June 2015
Development of the guideline (stage 3)	July – September 2015
Publication and post-publication activities (stage 4)	October 2015 – October 2016

REFERENCES

1. Whiting PF, Rutjes AW, Westwood ME, Mallett S: **A systematic review classifies sources of bias and variation in diagnostic test accuracy studies.** *J Clin Epidemiol* 2013, **66:**1093-1104.
2. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC: **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.
3. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, Bossuyt PM: **Reporting Diagnostic Accuracy Studies: Some Improvements after 10 Years of STARD.** *Radiology* 2014**:**141160.
4. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF: **CONSORT for reporting randomised trials in journal and conference abstracts.** *Lancet* 2008, **371:**281-283.
5. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, Gotzsche PC, Lasserson T, Tovey D: **PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts.** *PLoS Med* 2013, **10:**e1001419.
6. Korevaar DA, Cohen JF, Hooft L, Bossuyt PM: **Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies.** *J Clin Epidemiol* 2015 [Epub ahead of print].
7. Moher D, Schulz KF, Simera I, Altman DG: **Guidance for developers of health research reporting guidelines.** *PLoS Med* 2010, **7:**e1000217.
8. Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Hammick M, Moher D, Tilson J, Williams MT: **Protocol for development of the guideline for reporting evidence based practice educational interventions and teaching (GREET) statement.** *BMC Med Educ* 2013, **13:**9.
9. Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Hammick M, Moher D, Tilson JK, Williams MT: **A Delphi survey to determine how educational interventions for evidence-based practice should be reported: stage 2 of the development of a reporting guideline.** *BMC Med Educ* 2014, **14:**159.

10. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF: **CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration.** *PLoS Med* 2008, **5:**e20.
11. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM: **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.** *Ann Intern Med* 2011, **155:**529-536.
12. Deeks JJ, Altman DG: **Inadequate reporting of controlled trials as short reports.** *Lancet* 1998, **352:**1908.
13. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ: **More informative abstracts revisited.** *Ann Intern Med* 1990, **113:**69-76.
14. Kho ME, Eva KW, Cook DJ, Brouwers MC: **The Completeness of Reporting (CORE) index identifies important deficiencies in observational study conference abstracts.** *J Clin Epidemiol* 2008, **61:**1241-1249.
15. Brazzelli M, Lewis SC, Deeks JJ, Sandercock PA: **No evidence of bias in the process of publication of diagnostic accuracy studies in stroke submitted as abstracts.** *J Clin Epidemiol* 2009, **62:**425-430.
16. Estrada CA, Bloch RM, Antonacci D, Basnight LL, Patel SR, Patel SC, Wiese W: **Reporting and concordance of methodologic criteria between abstracts and articles in diagnostic test studies.** *J Gen Intern Med* 2000, **15:**183-187.
17. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32:**1008-1015.