Reporting Weaknesses in Conference Abstracts of Diagnostic Accuracy Studies in Ophthalmology

Daniël A. Korevaar, MD, **Jérémie F. Cohen, MD, PhD**, **Maurice W. J. de Ronde, MD**, **Gianni Virgili, MD**, **Kay Dickersin, PhD**, and **Patrick M. M. Bossuyt, PhD**

Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands (Korevaar, Cohen, de Ronde, Bossuyt); Inserm U1153, Center for Epidemiology and Statistics Sorbonne Paris Cité, Paris Descartes University, Paris, France (Cohen); Department of Translational Surgery and Medicine, Eye Clinic, University of Florence, Florence, Italy (Virgili); Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Dickersin).

Abstract

IMPORTANCE—Conference abstracts present information that helps clinicians and researchers to decide whether to attend a presentation. They also provide a source of unpublished research that could potentially be included in systematic reviews. We systematically assessed whether conference abstracts of studies that evaluated the accuracy of a diagnostic test were sufficiently informative.

OBSERVATIONS—We identified all abstracts describing work presented at the 2010 Annual Meeting of the Association for Research in Vision and Ophthalmology. Abstracts were eligible if they included a measure of diagnostic accuracy, such as sensitivity, specificity, or likelihood ratios. Two independent reviewers evaluated each abstract using a list of 21 items, selected from published guidance for adequate reporting. A total of 126 of 6310 abstracts presented were eligible. Only a minority reported inclusion criteria (5%), clinical setting (24%), patient sampling (10%), reference standard (48%), whether test readers were masked (7%), 2×2 tables (16%), and confidence intervals around accuracy estimates (16%). The mean number of items reported was 8.9 of 21 (SD, 2.1; range, 4-17).

CONCLUSIONS AND RELEVANCE—Crucial information about study methods and results is often missing in abstracts of diagnostic studies presented at the Association for Research in Vision

and Ophthalmology Annual Meeting, making it difficult to assess risk for bias and applicability to specific clinical settings.

Diagnostic accuracy studies evaluate how well a test distinguishes diseased from nondiseased individuals by comparing the results of the test under evaluation ("index test"), with the results of a reference (or "gold") standard. Deficiencies in study design can lead to biased accuracy estimates, suggesting a level of performance that can never be reached in clinical practice. In addition, because of variability in disease prevalence, patient characteristics, disease severity, and testing procedures, accuracy estimates may vary across studies evaluating the same test.[1] For example, in one Cochrane review, the sensitivity of optical coherence tomography in detecting clinically significant macular edema in patients with diabetic retinopathy ranged from 0.67 to 0.94 across included studies and specificity ranged from 0.61 to 0.97.[2]

Given these potential constraints, readers of diagnostic accuracy study reports should be able to judge whether the results could be biased and whether the study findings apply to their specific clinical practice or policy-making situation.[3,4]

Conference abstracts often are short reports of actual studies, presenting information that helps clinicians and researchers to decide whether to attend a presentation. They also provide a source of unpublished research that could potentially be included in systematic reviews.[5] These decisions should be based on an early appraisal of the risk for bias and applicability of the abstracted study. We systematically evaluated the informativeness of abstracts of diagnostic accuracy studies presented at the 2010 Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO).

Methods

The online abstract proceedings from ARVO were searched for diagnostic accuracy studies presented in 2010 (eTable 1 in the Supplement). One reviewer (D.A.K.) assessed identified abstracts for eligibility. Abstracts were included if they reported on the diagnostic accuracy of a test in humans and stated that they calculated 1 or more of the following accuracy measures: sensitivity, specificity, predictive values, likelihood ratios, area under the receiver operating characteristic curve, or total accuracy.

For each abstract, one reviewer (D.A.K.) extracted the research field, commercial relationships, support, study design, sample size, and word count (Table 1). Extraction was independently verified by a second reviewer (J.F.C. or M.W.J.dR.).

The informativeness of abstracts was evaluated using a previously published list of 21 items, selected from existing guidelines for adequate reporting (Table 2; eTable 2 in the Supplement).[6] The items focus on study identification, rationale, aims, design, methods for participant recruitment and testing, participant characteristics, estimates of accuracy, and discussion of findings. Two reviewers (D.A.K. and J.F.C./M.W.J.dR.) independently scored each abstract. Disagreements were solved through discussion.

Results

Of 6310 abstracts accepted at ARVO 2010, we identified 126 as reporting on diagnostic accuracy studies (eReferences in the Supplement). Abstract characteristics are provided in Table 1. The most common target condition was glaucoma (n = 51); corresponding studies mostly (n = 39) evaluated imaging of the retinal nerve fiber layer, other retina and choroid structures, or optic disc morphology. Ocular surface and corneal disease (keratoconus and dry eye) and common chorioretinal diseases (diabetic retinopathy and age-related macular degeneration) were targeted in 16 and 15 studies, respectively, followed by various types of uveitis and optic nerve diseases in 9 and 7 studies, respectively.

The reporting of individual items is presented in Table 2; examples of complete reporting per item are provided in eTable 3 in the Supplement. Several elements that are crucial when assessing risk for bias or applicability of the study findings were rarely reported: inclusion criteria (5%), clinical setting (24%), patient sampling (10%), reference standard (48%), masking of test readers (7%), 2×2 tables (16%), and confidence intervals around accuracy estimates (16%). None of the abstracts reported all of these items. Reporting was better for other crucial elements: study design (87%), test under evaluation (100%), number of participants (82%), and disease prevalence (80%).

On average, the abstracts reported 8.9 of the 21 items (SD, 2.1; range, 4-17). Twenty-four abstracts (19%) reported more than half of the items (Figure). The mean number of reported items was significantly lower in abstracts of case-control studies compared with cohort studies ($P = .001$) and in abstracts with sample sizes (number of eyes) below the median ($P = .03$) (Table 1).

Discussion

The informativeness of abstracts of diagnostic accuracy studies presented at the 2010 ARVO Annual Meeting was suboptimal. Several key elements of study methods and results were rarely reported, making it difficult for clinicians and researchers to evaluate method quality.

Differences in patient characteristics and disease severity are known sources of variability in accuracy estimates, and nonconsecutive sampling of patients can lead to bias.[1,4] There fore, readers want to know where and how patients were recruited,[3] yet less than a quarter of abstracts reported inclusion criteria, clinical setting, and sampling methods.

Risk for bias and applicability largely depend on the appropriateness of the reference standard.[4] However, the reference standard was not reported in half of the abstracts. Agreement between 2 tests is likely to increase if the reader of one test is aware of the results of the other test[1,4]; however, information about masking was available in only 7%.

About half of all conference abstracts are never published in full.[5] It is only possible to include the results of a conference abstract in a meta-analysis if the number of true-positive, true-negative, false-positive, and false-negative test results are provided; however, 2×2 tables were only available in 16%. Although it is widely recognized that point estimates of

diagnostic accuracy should be interpreted with measures of uncertainty, confidence intervals were reported in 16%.

Other crucial elements were more frequently provided. The study design, reported by 87%, is important because case-control studies produce inflated accuracy estimates owing to the extreme contrast between participants with and without the disease.[1,7] Diagnostic accuracy varies with disease prevalence, an important determinant of the applicability of study findings, and reported by 80%.

Suboptimal reporting in conference abstracts is not only a problem for diagnostic accuracy studies.[8] A previous evaluation of the content of abstracts of randomized trials presented at the ARVO Annual Meeting also found important study design information frequently unreported.[9] However, the authors concluded that missing information was often available in the corresponding ClinicalTrials.gov record. Because diagnostic accuracy studies are rarely registered,[10] complete reporting of conference abstracts is even more critical for these studies.

Using the same list of 21 items, we previously evaluated abstracts of diagnostic accuracy studies published in high-impact journals.[6] The overall mean number of items reported there was 10.1; crucial items about design and results were similarly lacking. One previous study assessed elements of reporting in conference abstracts of diagnostic accuracy studies in stroke research.[11] In line with our findings, 35% reported whether the data collection was prospective or retrospective, 24% reported on masking, and 11% reported on test reproducibility. Incomplete reporting is not only a problem for abstracts. Five previous reviews evaluated the reporting quality of full-study reports of ophthalmologic diagnostic accuracy studies, all of them pointing to important shortcomings.[12]

Conclusions

Crucial study information is often missing in abstracts of diagnostic accuracy studies presented at the ARVO Annual Meeting. Suboptimal reporting impedes the identification of high-quality studies from which reliable conclusions can be drawn. This is a major obstacle to evidence synthesis and an important source of avoidable research waste.[13]

Our list of 21 items is not a reporting checklist; we are aware that word count restrictions make it impossible to report all items in an abstract, and some items are more important than others. Reporting guidelines have been developed for abstracts of randomized trials and systematic reviews,[8,14] and a similar initiative is currently under way for diagnostic abstracts.[15] The scientific community should encourage informative reporting, not only for full-study reports, but also for conference abstracts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Whiting PF, Rutjes AW, Westwood ME, Mallett S. QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol. 2013; 66(10):1093–1104. [PubMed: 23958378]

2. Virgili G, Menchini F, Casazza G, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. Cochrane Database Syst Rev. 2015; 1:CD008081. [PubMed: 25564068]

3. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ. 2003; 326(7379):41–44. [PubMed: 12511463]

4. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155(8):529–536. [PubMed: 22007046]

5. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev. 2007; (2):MR000005. [PubMed: 17443628]

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; 68(6): 708–715. [PubMed: 25703213]

7. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005; 51(8):1335–1341. [PubMed: 15961549]

8. Hopewell S, Clarke M, Moher D, et al. CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS Med. 2008; 5(1):e20. [PubMed: 18215107]

9. Scherer RW, Huynh L, Ervin AM, Taylor J, Dickersin K. ClinicalTrials.gov registration can supplement information in abstracts for systematic reviews: a comparison study. BMC Med Res Methodol. 2013; 13:79. [PubMed: 23773868]

10. Korevaar DA, Bossuyt PM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. BMJ Open. 2014; 4(1):e004596.

11. 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(4):425–430. [PubMed: 19013759]

12. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med. 2014; 19(2):47–54. [PubMed: 24368333]

13. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014; 383(9913):267–276. [PubMed: 24411647]

14. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts Group. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. PLoS Med. 2013; 10(4):e1001419. [PubMed: 23585737]

15. Cohen, JF.; Korevaar, D.; Hooft, L.; Reitsma, JB.; Bossuyt, PM. [August 10, 2015] Development of STARD for abstracts: essential items in reporting diagnostic accuracy studies in journal or conference abstracts. http://www.equator-network.org/wp-content/uploads/2009/02/STARD-for-Abstracts-protocol.pdf. 2015.

At a Glance

- Understanding the informative value of ophthalmology abstracts might lead to improved content in the future.

- Abstracts of diagnostic accuracy studies presented at the 2010 Annual Meeting of the Association for Research in Vision and Ophthalmology were evaluated.

- A minority reported inclusion criteria (5%), clinical setting (24%), patient sampling (10%), the gold standard used (48%), and masking (7%).

- Reporting was better for study design (87%), the test under evaluation (100%), number of participants (82%), and disease prevalence (80%).

- This study exemplified how deficiencies in abstracts may make it difficult to assess risk for bias and applicability to specific clinical settings.
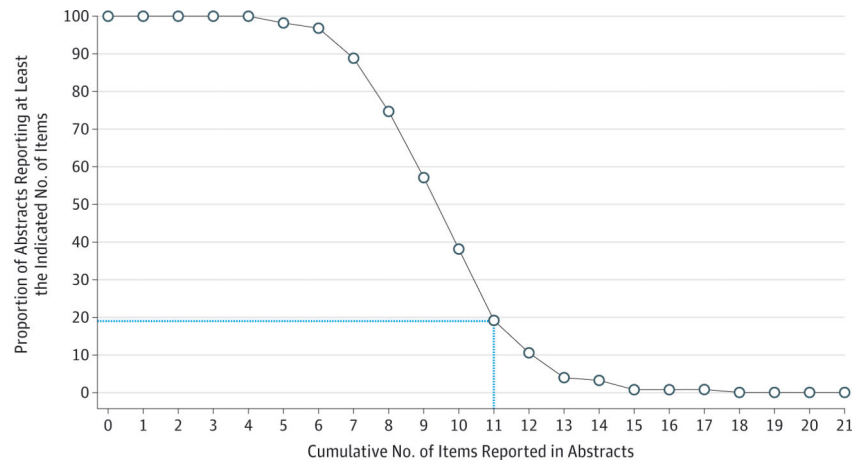
Figure. Proportion of Diagnostic Abstracts (N=126) That Reported at Least the Indicated Number of Items on the 21-Item List

The blue dotted line indicates the percentage of abstracts reporting more than half of the evaluated items.

Table 1

Mean Number of Items Reported Among Diagnostic Abstracts (N=126), Stratified by Study Characteristics

Study Characteristic	No. (%)	No. of Items Reported, Mean (SD)	P Value[a]
Research field			
Glaucoma	51 (41)	9.1 (1.6)	.35
Other than glaucoma	75 (59)	8.8 (2.4)	
Ocular surface and corneal diseases	16 (13)	NA	
Common chorioretinal diseases	15 (12)	NA	
Various types of uveitis	9 (7)	NA	
Optic nerve diseases	7 (6)	NA	
Other	28 (22)	NA	
Commercial relationships			
1 Author	44 (35)	8.9 (2.0)	.85
No author	82 (65)	9.0 (2.2)	
Support			
Industry support	12 (10)	8.4 (1.4)	.58
No industry support	114 (90)	9.0 (2.2)	
Study design[b]			
Cohort	38 (35)	10.1 (2.5)	.001
Case-control	72 (66)	8.6 (1.5)	
No. of patients in sample, median (IQR)[c]	100 (50-160)		
<100	50 (49)	9.0 (2.4)	.26
100	53 (51)	9.5 (1.8)	
No. of eyes in sample, median (IQR)[d]	136 (55-219)		
<136	33 (49)	8.4 (2.0)	.03
136	34 (51)	9.4 (1.9)	
Word count, median (IQR)[e]	301 (255-327)		
<301	61 (49)	8.8 (2.0)	.40
301	65 (51)	9.1 (2.3)	

Abbreviations: IQR, interquartile range; NA, not applicable.

[a]Mean number of items reported across subgroups was compared using the *t* test.

[b]Study design was unclear for 16 abstracts.

[c]Sample size (number of patients) was unclear for 23 abstracts.

[d]Sample size (number of eyes) was unclear or NA for 59 abstracts.

[e]Abstract word count excluding title, affiliations, commercial relationships, support, references, keywords, tables, and figures.

Table 2

Items Reported in Diagnostic Abstracts (N=126)

Item	No. (%)
Title	
Identify the article as a study of diagnostic accuracy in title	57 (45)
Background and aims	
Rationale for study/background	34 (27)
Research question/aims/objectives	103 (82)
Methods	
Study population, at least 1 of the following	36 (29)
Inclusion/exclusion criteria	6 (5)
Clinical setting	30 (24)
No. of centers	25 (20)
Study location	18 (14)
Recruitment dates	15 (12)
Patient sampling, consecutive vs random sample	12 (10)
Data collection, prospective vs retrospective	27 (21)
Study design, case-control vs cohort	110 (87)
Reference standard	61 (48)
Information on the index test under evaluation, at least 1 of the following	126 (100)
Index test	126 (100)
Technical specifications and/or commercial name	101 (80)
Cutoffs and/or categories of results of index test	40 (32)
Whether test readers were masked, at least 1 of the following	9 (7)
When interpreting the index test	6 (5)
When interpreting the reference standard	5 (4)
Results	
Study participants, at least 1 of the following	107 (85)
No. of participants	103 (82)
Age of participants	27 (21)
Sex of participants	7 (6)
Information on indeterminate results/missing values	15 (12)
Disease prevalence	101 (80)
2×2 Tables, No. of true- and false-positive and -negative test results	20 (16)
Estimates of diagnostic accuracy, at least 1 of the following	122 (97)
Sensitivity and/or specificity	84 (67)
Negative and/or positive predictive value	14 (11)
Negative and/or positive likelihood ratio	1 (1)
Area under the ROC curve/C statistic	56 (44)
Diagnostic odds ratio	1 (1)

Item	No. (%)
Accuracy	6 (5)
95% CIs around estimates of diagnostic accuracy	20 (16)
Reproducibility of the results of the index test under evaluation	6 (5)
Discussion/conclusion	
Discussion of diagnostic accuracy results	120 (95)
Implications for future research	23 (18)
Limitations of study	1 (1)

Abbreviation: ROC, receiver operating characteristic.