Diagnostic accuracy studies need more informative abstracts

Jérémie F. Cohen [1,2] · Daniël A. Korevaar [3] · Patrick M. Bossuyt [4]

In the era of evidence-based medicine, clinicians, scientists, and policymakers have to screen large amounts of information in short periods of time to stay up-to-date. Journal abstracts play a crucial role in this process. Based on the abstract, readers commonly decide whether they should invest time in obtaining and reading the full text. Sometimes, decisions regarding the adoption of interventions and medical tests may be made on abstracts alone, for example, when full texts are not available because of a paywall.

The abstract should contain the core elements of a study so that readers can quickly identify whether it is relevant to their field of practice. Which information is critical to report in an abstract varies between study types. For comparative studies of interventions, it means that minimal information regarding patients, intervention, control, and outcomes (PICO) should be reported in the abstract. For diagnostic accuracy studies, the authors should at least disclose in the abstract where, in whom, and how the test was evaluated. By diagnostic accuracy studies, we mean studies in which the results of the test(s) under evaluation are compared with those obtained with the reference standard in a series of patients suspected of having a given condition or disease. The typical outcomes of diagnostic accuracy studies are sensitivity and specificity, but sometimes, authors may prefer predictive values or likelihood ratios [1].

The accuracy of a diagnostic test is not a fixed property; sensitivity and specificity vary with patient population and setting [2, 3]. Also, shortcomings in study design may generate biased estimates of diagnostic accuracy. For example, it was shown that studies including healthy controls systematically overestimate sensitivity and specificity [4]. Abstracts of diagnostic accuracy studies should, therefore, report essential information needed to assess the risk of bias from limitations in study design and conduct, the actual study findings, and potential concerns about the applicability of the findings to readers' clinical questions.

In the last issue of the journal, Azrad and colleagues report on a study in which they compared the diagnostic accuracy of two rapid tests for strep throat, with throat culture as the clinical reference standard [5]. Based on estimates of sensitivity and specificity, they conclude that "Both tests have a good diagnostic performance." Independent of the overall quality and relevance of the Azrad study, their abstract might be considered insufficiently informative: eligibility criteria (including whether the study focused on children, adults, or both), setting characteristics (including the nature and number of centers involved), key figures (including the number of participants with and without strep throat), and 95% confidence intervals around accuracy estimates are not reported. We also do not know from the abstract whether the study has been registered in a publicly accessible trial registry before its initiation.

We believe Azrad and colleagues' abstract illustrates common deficiencies in the reporting of abstracts of diagnostic accuracy studies. In a sample of journal and conference abstracts of diagnostic accuracy studies, we found that eligibility criteria, study setting, and confidence intervals around accuracy estimates were reported in less than half of abstracts [6, 7].

✉ Jérémie F. Cohen
jeremie.cohen@inserm.fr

Daniël A. Korevaar
d.a.korevaar@amc.uva.nl

Patrick M. Bossuyt
p.m.bossuyt@amc.uva.nl

1 Department of General Pediatrics and Pediatric Infectious Diseases, Necker-Enfants Malades Hospital, APHP, Paris Descartes University, 149 rue de Sèvres, 75015 Paris, France

2 Inserm UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Research Center for Epidemiology and Biostatistics Sorbonne Paris Cité (CRESS), Paris Descartes University, 53 avenue de l'Observatoire, 75014 Paris, France

3 Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

4 Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Why does this matter? On average, the sensitivity and specificity of rapid tests for strep throat are around 85% and 95% [8]. However, the accuracy of rapid tests is well known to vary from study to study because of patient characteristics, such as age and Centor criteria [9, 10]. The accuracy of rapid tests also depends on the person performing the test. For example, in a study evaluating sources of variability in the accuracy of a rapid test for strep throat, sensitivity ranged from 56 to 96% ($p = 0.01$) across physicians, and was higher for physicians with hospital-based clinical activity in addition to office-based practice (91% vs. 78%, $p < 0.001$) [9]. Sensitivity may also be higher in studies with higher disease prevalence [11]. Confidence intervals matter because presenting point estimates obtained from small studies, without expressing the statistical imprecision, might suggest a too precise reflection of the "true" accuracy of the test. Azrad and colleagues' full text shows that, for both rapid tests, the 95% confidence interval of sensitivity ranges from 60.9 to 91.1%, around a point estimate of 80%. Prospective registration, for example, in Clinicaltrials.gov, allows the timely detection of ongoing and completed studies and enables the investigation of deviations from the original study protocol [12]. Suboptimal reporting of study findings and unbalanced abstracts may also make diagnostic tests seem more favorable than they are, a phenomenon sometimes referred to as "spin" [13].

Incomplete reporting is highly prevalent in the peer-reviewed literature [6, 14]. To improve this situation, a large number of reporting guidelines have been developed, for different types of research. Many journals now explicitly invite researchers to adhere to such guidelines. For example, in 2014, 63% of high-impact journals mentioned CONSORT (Consolidated Standards Of Reporting Trials), the main reporting guideline for randomized trials, in their instructions to authors [15].

CONSORT has many extensions, including one providing clear guidance regarding the content of abstracts of reports of randomized trials [16]. Similarly, to help authors, reviewers, and editors of diagnostic accuracy studies, an international group of experts has recently developed STARD for Abstracts [17]. This checklist contains 11 items that can be included in any abstract of a diagnostic accuracy study (checklist available at www.equator-network.org/reporting-guidelines/stard-abstracts). STARD for Abstracts also comes with examples of good reporting and template text that can be used by authors when writing their abstracts. The STARD for Abstracts reporting guideline is an extension of STARD 2015 (Standards for Reporting Diagnostic Accuracy), the updated reporting guideline for full study reports of diagnostic accuracy studies [18, 19].

There is evidence that completeness of reporting improved in 10 years after the launch of STARD [20]. We believe the use of STARD for Abstracts can further improve the informativeness of journal and conference abstracts of diagnostic accuracy studies, thereby helping readers and reducing avoidable research waste [14]. Implementation of reporting guidelines at the journal may also contribute to enhance the usability of reports published in the biomedical literature.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflicts of interest.

Ethical approval   Not applicable.

Informed consent   Not applicable.

References

1. Cohen JF, Chalumeau M, Bossuyt PM (2017) Clinicians should not be forced to use likelihood ratios when comparing tests. Eur J Clin Microbiol Infect Dis 36(1):195–196
2. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH et al (1999) Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 282(11):1061–1066
3. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM (2006) Evidence of bias and variation in diagnostic accuracy studies. CMAJ 174(4):469–476
4. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM (2005) Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem 51(8):1335–1341
5. Azrad M, Danilov E, Goshen S, Nitzan O, Peretz A. Detection of group a Streptococcus in pharyngitis by two rapid tests: comparison of the BD Veritor and the QuikRead go(R) Strep A. Eur J Clin Microbiol Infect Dis. https://doi.org/10.1007/s10096-019-03527-w
6. Korevaar DA, Cohen JF, Hooft L, Bossuyt PM (2015) Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. J Clin Epidemiol 68(6):708–715
7. Korevaar DA, Cohen JF, de Ronde MW, Virgili G, Dickersin K, Bossuyt PM (2015) Reporting weaknesses in conference abstracts of diagnostic accuracy studies in ophthalmology. JAMA Ophthalmol 133(12):1464–1467
8. Cohen JF, Bertille N, Cohen R, Chalumeau M (2016) Rapid antigen detection test for group A streptococcus in children with pharyngitis. Cochrane Database Syst Rev (7):CD010502
9. Cohen JF, Chalumeau M, Levy C, Bidet P, Benani M, Koskas M et al (2013) Effect of clinical spectrum, inoculum size and physician characteristics on sensitivity of a rapid antigen detection test for group A streptococcal pharyngitis. Eur J Clin Microbiol Infect Dis 32(6):787–793
10. Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A et al (2012) Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. PLoS One 7(6):e39085
11. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM (2013) Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 185(11):E537–E544
12. Korevaar DA, Hooft L, Askie LM, Barbour V, Faure H, Gatsonis CA et al (2017) Facilitating prospective registration of diagnostic accuracy studies: a STARD initiative. Clin Chem 63(8):1331–1341
13. Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeflang MM (2013) Overinterpretation and misreporting of diagnostic accuracy studies: evidence of "spin". Radiology 267(2):581–588

14. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S et al (2014) Reducing waste from incomplete or unusable reports of biomedical research. Lancet 383(9913):267–276

15. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF (2016) Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "instructions to authors" in 2014. Trials 17(1):301

16. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG et al (2008) CONSORT for reporting randomised trials in journal and conference abstracts. Lancet 371(9609):281–283

17. Cohen JF, Korevaar DA, Gatsonis CA, Glasziou PP, Hooft L, Moher D et al (2017) STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. BMJ (Clin Res ed.) 358:j3751

18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L et al (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ (Clin Res ed.) 351: h5527

19. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft A et al (2016) STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 6: e012799

20. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N et al (2014) Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. Radiology:141160