Guidance was developed on how to write a plain language summary for diagnostic test accuracy reviews

Whiting, Penny F; Leeflang, Mariska M G; de Salis, Isabel ; Mustafa, Reem A. ; Santesso, Nancy ; Gopalakrishna, Gowri ; Cooney, Geraldine ; Jesper, Emily ; Thomas, Joanne ; Davenport, Clare

DOI:
10.1016/j.jclinepi.2018.07.008

Document Version
Peer reviewed version

Citation for published version (Harvard):
Whiting, PF, Leeflang, MMG, de Salis, I, Mustafa, RA, Santesso, N, Gopalakrishna, G, Cooney, G, Jesper, E, Thomas, J & Davenport, C 2018, 'Guidance was developed on how to write a plain language summary for diagnostic test accuracy reviews', *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2018.07.008

How to write a plain language summary for a diagnostic test accuracy review

Penny Whiting,[1,2] Mariska Leeflang,[3] Isabel de Salis,[1] Reem A. Mustafa,[4, 5] Nancy Santesso,[5] Gowri Gopalakrishna,[3] Geraldine Cooney,[2] Emily Jesper,[7] Joanne Thomas,[7] Clare Davenport[8]

[1] Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom

[2] NIHR CHLARC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

[3] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

[4] Division of Nephrology and Hypertension, and Director, Outcomes and Implementation Research | University of Kansas Medical Center , Kansas City, Kansas, United States

[5] Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada

[6]Sense about Science, 14a Clerkenwell Green, London, United Kingdom

[7]'Institute of Applied Health Research', University of Birmingham, Birmingham, United Kingdom

Corresponding author: Dr Penny Whiting, NIHR CHLARC West, 9th Floor, Whitefriars, Lewins Mead, Bristol; penny.whiting@bristol.ac.uk

Declarations of interest: none

Role of the funding sources:

Financial support for this study was provided in part by a Cochrane's Methods Innovation Fund. PW (time) was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) West at University Hospitals Bristol NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social care. The funding agreement ensured the

authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Abstract [132 words]

Objective

To develop guidance for authors of diagnostic test accuracy (DTA) reviews to help them write a plain language summary (PLS) of the results of their review.

Study design and setting

We used a combination of focus groups, user testing and a web based survey. Participants included patient representatives, media representatives and health professionals.

Results

We present step-by-step guidance for authors of DTA reviews for writing a PLS. This guidance is illustrated with examples of reader-tested sentences, explanations and a figure.

Conclusion

We hope this guidance will allow reviewers to present the findings of DTA reviews so that it is easier for readers to understand the results and conclusions. This will increase the accessibility of these reviews for various audiences.

Key words: Diagnostic test accuracy, systematic review, plain language summary, sensitivity and specificity, natural frequencies

What is new?

- Understanding and application of test accuracy evidence is challenging
- Improving the accessibility of test accuracy evidence has the potential to positively impact test use.
- This paper introduces new guidance for reviewers on how to write plain language summaries of diagnostic test accuracy reviews
- The aim of this guidance is to help reviewers to present the results of DTA reviews to make them easier for readers to understand

Introduction

A plain language summary (PLS) is an easy to read summary of a systematic review and should provide rapid access to the content of the review.[1] Just like the abstract of a manuscript, PLS are generally made freely available on the internet, so will often be read as stand-alone documents.

A clear PLS is essential to ensure that systematic reviews are accessible to users who are not familiar with the more technical content of a review. Complexity of methods and understanding of diagnostic accuracy measures is likely to increase the diversity of the audience for PLS for diagnostic research. Thus, users of a PLS may not be limited to the public but may also include health care professionals, policy makers and the media.

Explaining the results of a Diagnostic Test Accuracy (DTA) review in plain language presents particular challenges. The review methodology and terminology are less familiar than reviews of interventions.[2] Commonly used measures of test accuracy such as sensitivity, specificity, positive and negative predictive values and likelihood ratios are poorly understood by health professionals.[3] Research has shown that readers familiar with systematic review methods have difficulties understanding DTA reviews and that familiarity with intervention reviews may even be a disadvantage.[4] DTA reviews are concerned with evaluating test accuracy and do not directly evaluate whether introduction of a new test will result in better outcomes for patients. The potential benefits of improved test accuracy will only be realised if introduction of a new test results in a change in diagnostic yield, in other words revision of a diagnosis leading to an appropriate change in patient management. This requires that health professionals have access to a new test, the skill to conduct and interpret a new test, confidence in the test result and following this that appropriate treatment is initiated and is effective. [5]

Reporting test accuracy using natural frequencies and visual aids rather than using probabilistic language may facilitate improved understanding and better estimation of the post-test probability of disease.[3] In addition, DTA reviews are characterised by a large degree of heterogeneity in results across studies, the reason for this variation is not always clear and explaining this to users is difficult.[4] Sources of bias in DTA studies differ from those of intervention studies, and implications of the impact of bias are not always clearly understood. [6]

Here, we present step-by-step guidance for authors of DTA reviews for writing a PLS. This guidance is illustrated with examples of reader-tested sentences, explanations and a figure.

Methods

We prepared this guidance based on the findings of research funded by the Cochrane Collaboration and drawing on the PLS guidance for Cochrane Intervention Reviews.[1] Although the template and guidance were developed primarily for Cochrane DTA reviews they are equally relevant to any DTA review.

We used a mixed methods approach consisting of focus groups, user testing, web-based surveys and a public engagement event to develop this guidance (Figure 1).

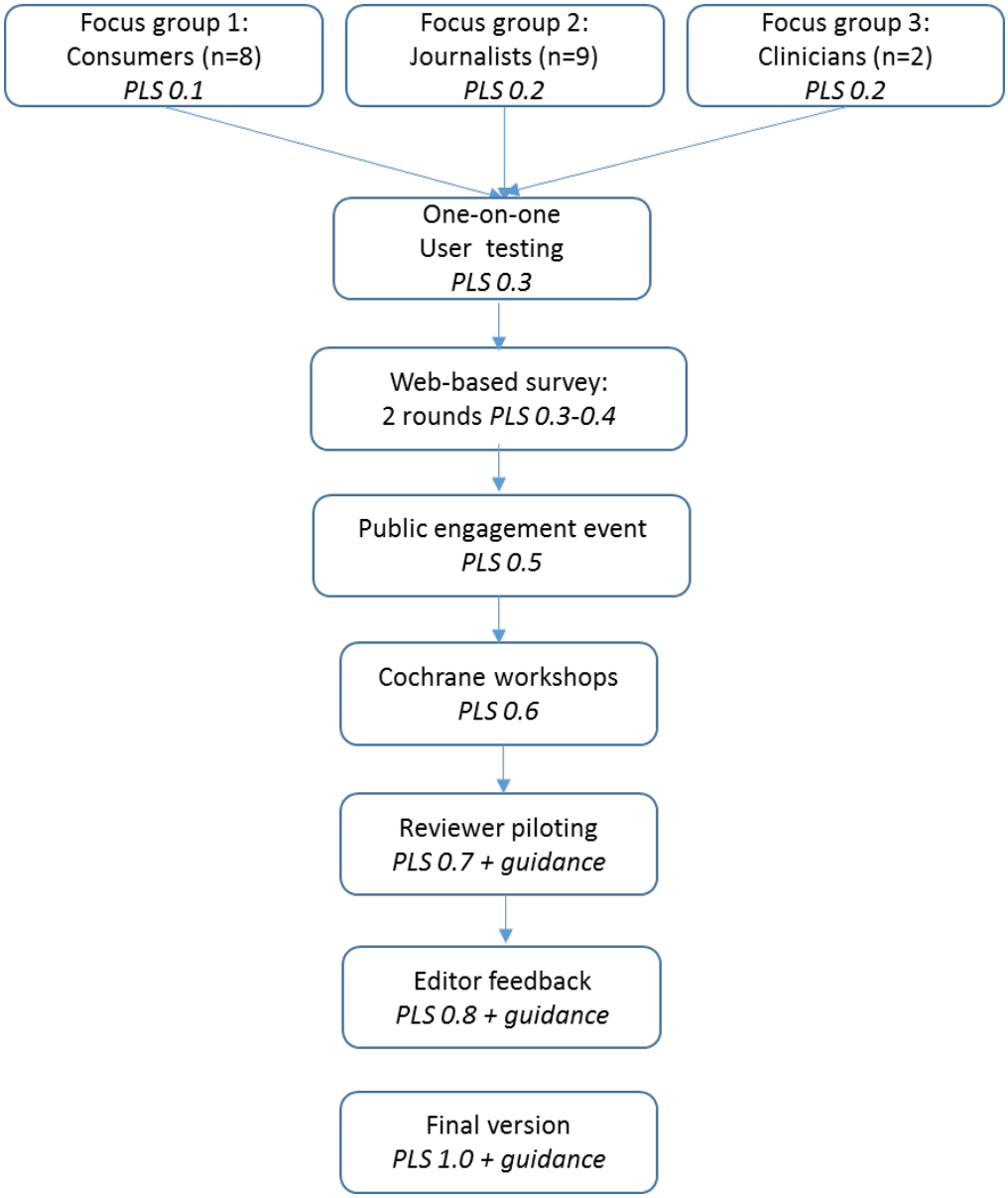


Figure 1: Outline of process used to develop PLS guidance

Initial focus groups were conducted with a range of potential end-users including one with consumers (8 participants), one with journalists (9 participants) and one with clinicians (2 participants). During the focus groups we presented two example PLS for discussion – an existing PLS from a DTA review and a PLS that we rewrote based on guidance on how to structure a PLS for a Cochrane Intervention review,[1] with some modifications to fit the template to DTA reviews (PLS 0.1). Some minor corrections to wording of the PLS were made following focus group 1 (PLS 0.2). We also included several alternative methods for presenting the numerical results of the review. Participants were asked about their general views on the two PLS, what they liked and disliked about each, and how the two compared. We then asked them about their views on how numerical results should be presented and which of the four alternative suggestions presented they preferred. There was a clear preference for our new suggested structure and to include a figure, there were also a number of suggestions for improvements. Based on the results of the focus group we produced a revised PLS with substantial changes to wording and headings and inclusion of a figure to summarise numerical results. [7] This updated PLS (PLS 0.3) was used for the next stage of the development process - one-on-one user testing with potential stake holders: four clinicians, one journalist, one commissioner, one review author and one patient representative. All supported the changes made following the focus groups with some additional changes suggested to the wording of some sections.

Concurrently with the user testing, we ran a web-based survey to gain feedback from a wider group of participants on the same version of the PLS that was used for the user testing (PLS 0.3). This was completed by 67 respondents including media representatives, methodologists, systematic review authors, health professionals, and patient representatives. There was strong support for the proposed structure and numerical presentation of results, although suggestions were made to improve the wording of the PLS, including the title and section headings. A refined version was produced (PLS 0.4) and was considered as part of the second round of the survey. Minor wording changes were made to this following the second round of the survey (PLS 0.5) and this version was used for a public engagement event where we presented the example PLS to a group of 16 to 18 year old school students. Further minor wording changes were suggested as well as some changes to the figure to make it easier to read. They also suggested that it would be helpful to include a narrative summary of the numerical results in addition to the graphical display. We modified the PLS to take this feedback into account and shared this version for further feedback during workshops at Cochrane Colloquia (PLS 0.6). Following this we finalised the PLS (PLS 0.7) and produced a guidance document to accompany it. This version together with the guidance was piloted by a small number of reviewers and then shared with Cochrane DTA Editors following further minor adaptions (PLS 0.8).

All were positive about the PLS structure and guidance with only minor changes needed to finalise the guidance and example PLS included here.

Our research generated the following sub-headings for structuring a PLS. Below we provide a description of the content and suggested wording for each sub-heading, illustrated using an example PLS for a systematic review of rapid tests for diagnosing group A streptococcus infection in children with sore throat.[8] The complete PLS for this review is provided as a web appendix (Web Appendix 1). An additional complete PLS example based on a review of the IQCODE questionnaire to diagnose dementia in hospital[9] is also provided as a web appendix (Web Appendix 2).

Review title

If the review title is difficult to understand, for instance if it includes technical terms or jargon, consider re-writing it in plain language. As a minimum, the review title should contain information about the following three key elements:

The *test or tests* being studied (index tests). It is important to ensure the type of test being studied is clear (e.g. questionnaire, a blood test, a swab, a urine test or some form of medical imaging). For example, describing the index test as a 'rapid' test with no further information does not convey the implications of having the test for an individual.

The *condition* that the test is designed to detect (the target condition).

The *people* who will receive the test (for example adults, children, people with certain symptoms such as sore throat or low back pain). It may also be important to include any restrictions on the healthcare setting where the test will be applied; for example, if the test will only be used in hospital settings and not the community.

Example: "How accurate are rapid swab tests for strep throat in children?"

Why is improving [….] diagnosis important?

This subheading should include information about the target condition and how use of the index test might benefit individuals suspected of having the target condition. For example, the index test may be more accurate, may provide quicker results, or may be more accessible (less costly, require less expertise) than tests currently in use. A brief description of the downstream consequences of

testing should be included. It is helpful to introduce the concept of test errors - “false positive” and “false negative” at this early stage.

Our research to develop this guidance demonstrated that the downstream consequences of test errors were considered particularly important by potential users. A sentence on the benefits of making a correct diagnosis: a true positive (index test positive and target condition present) and a true negative (index test negative and target condition not present) test results is therefore helpful to include here.

- What are the consequences of a false positive result (index test positive but target condition not present, i.e. incorrectly labelling individuals as having the condition when they don’t)?
- What are the consequences of a false negative result (index test negative but target condition present, i.e. missing the diagnosis of the condition in an individual who has the condition)?

Example: ***Why is improving the diagnosis of bacterial infection diagnosis?***

Sore throat is very common in children. It can be caused by viruses or bacteria. Antibiotic treatment is only useful for sore throat caused by bacteria, which is usually caused by group A streptococcus ('strep throat'). Not recognizing bacterial infection when it is present (a false negative test result) may result in delayed recovery and an increased risk of infecting others. It may also result in rare but serious complications such as abscesses in the throat, bacterial infection of the sinuses and ears, and rheumatic fever. An incorrect diagnosis of bacterial infection (a false positive test result) may mean that children are given antibiotics when there is no benefit to be gained.

What is the aim of this review?

The aim of the review should be stated as concisely and simply as possible. Our research illustrated that users do not always understand that the results of a PLS come from a systematic review rather than a single study. Some also wrongly assume that the review authors have carried out the studies included in the review themselves. We therefore suggest using an introductory sentence such as:

“The aim of this review was to find out how accurate [….]. Researchers included [X#] studies to answer this question.”

Example: ***What is the aim of this review?***

The aim of this review was to find out how accurate rapid tests are for diagnosing bacterial infection in children with sore throat. Researchers included 98 studies to answer this question.

What was studied in the review?

Give a brief description of the review topic considering the following questions:

- What was the index test(s) addressed in the review? Give enough information for readers to judge whether the test(s) being studied is relevant to them, for example where in the clinical pathway is the test likely to be applied?
- "What is the role of the index test (e.g. triage, add-on, or replacement test) [10] Authors should avoid using these technical terms and instead describe how the index test would be placed in the current testing pathway.
- If there is more than one index test included in a review the PLS should also explain how these tests differFor example, one test may be quicker to give results or easier to perform, tests may be produced by a different manufacturer or require different processing techniques, one test may be blood test and another a swab test
- Our research demonstrated that the presence of the target condition may not be considered a "positive" outcome and so the term "positive" test result should be avoided. Instead describe the test result that indicates if the target condition is present.

Example: ***What was studied in the review?***

Two types of rapid tests were studied. These use different biochemical methods to identify the bacterial infection. Rapid tests require just a simple throat swab from the patient. This gives an immediate result allowing clinicians to decide whether to prescribe antibiotics. This is an advantage compared to conventional laboratory tests which take 48 hours to give a result.

What are the main results in this review?

Describing the included studies

In this section the number of included studies and total number of participants should briefly be described. To clarify that the number of participants applies to the sum total of participants across included studies it is helpful to structure this sentence as follows:

"The review included [x#] relevant studies with a total of [x#] participants."

Presenting information on test accuracy

We suggest presenting the summary accuracy data using natural frequencies based on a hypothetical cohort of 1000 patients.[7]

Presentation of sensitivity and specificity are the most commonly used metric in DTA reviews.[11] The tables below can be used to derive natural frequencies from summary sensitivities and specificities in a DTA review. Online tools are also available that will perform these calculations, for example GRADEPro.[12]

The numbers to populate this table can be calculated by taking summary estimates of sensitivity, specificity and prevalence (p) from the systematic review as follows:

Table 1: How to transform sensitivity, specificity and prevalence to natural frequencies

	Disease Present	Disease Absent	Total
Test Positive	TP = x*sens	FP = y – (y*spec)	TP+FP
Test Negative	FN = x – (x*sens)	TN = y*spec	FN+TN
	x=1000 * p	y=1000 – (1000*p)	1000

Example: Prevalence of disease = 30% (p=0.3), sensitivity = 85.6% (sens = 0.856), specificity = 95.4% (spec = 0.954)

Table 2: Worked example showing how to transform sensitivity, specificity and prevalence to natural frequencies

	Disease Present	Disease Absent	Total
Test Positive	300*0.856=257	700-(700*0.954)=32	TP+FP=289
Test Negative	300-(300*0.856)=43	700*0.954=668	FN+TN=711
	1000 *0.3=300	1000 – 300=700	1000

An estimate of the prevalence should be taken from the systematic review, either the mean or median prevalence across studies, unless there is a good rationale for taking an alternative estimate.

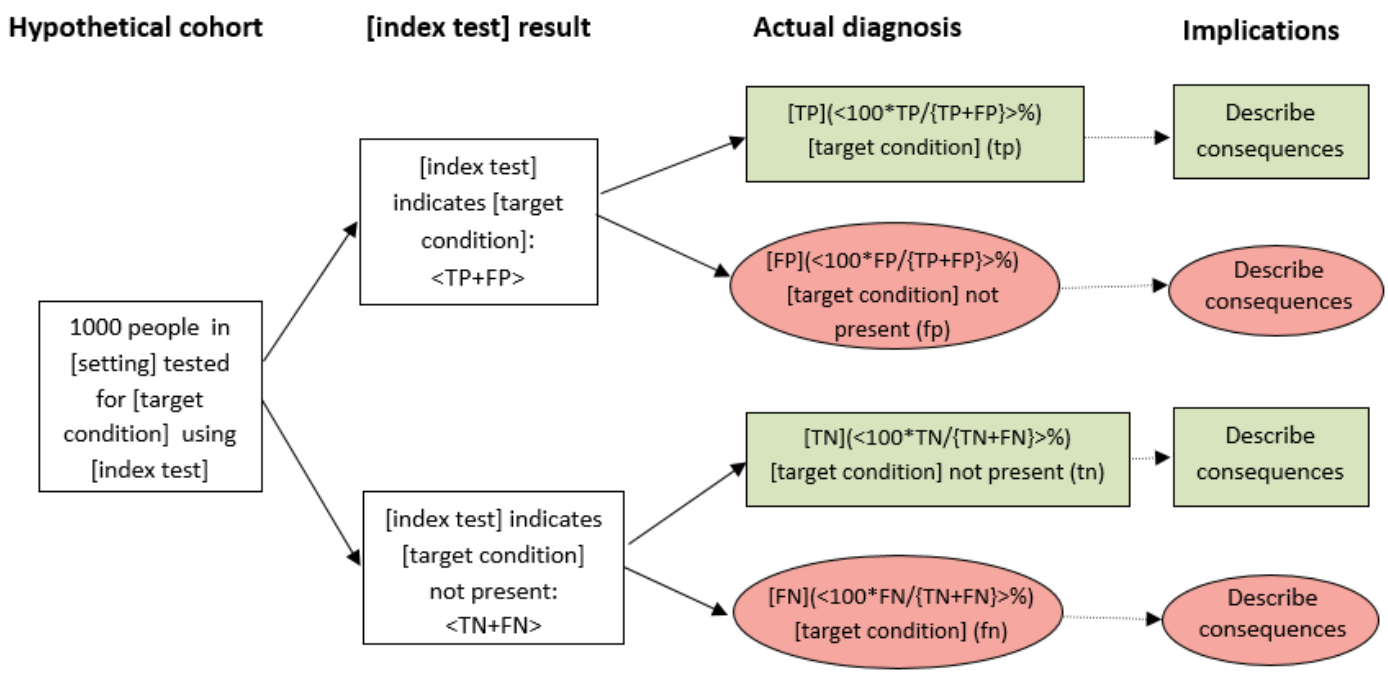
Suggested text for presenting accuracy data:

The results of these studies indicate that in theory, if the [index test] were to be used in [setting] in a group of 1000 people where [x (p%)] have [target condition]:

- *An estimated [TP + FP] would have an [index test] result indicating [target condition] is present and of these[FP](<100*FP/{TP+FP}>) wouldl be incorrectly classified as having the [target condition]*
- *Of the [TN + FN] people with a result indicating that [target condition] is not present, [FN](<100*FN/{TN+FN}>%) would be incorrectly classified as not having [target condition]*

We propose a flow diagram to summarise these results. This can be populated from the tables above combined with the information on implications that is described in the sub-heading *"Why is improving [….] diagnosis important?".*

Note that in the formulas above and in the figure template:
Text in square brackets, [ ], is replaced with text or numbers in the boxes in the final figure.
Text in Angle brackets, <> indicate expressions to be calculated.
Capitalised TP, FP, TN, FN indicate variables to be replaced with numbers.
Lower case tp, fp, tn, fn are shown unchanged in the final figure.



tp: true positive – test is positive (indicates [target condition]) and patient has [target condition]
fp: false positive – test is positive (indicates [target condition]) but patient does not have [target condition]
tn: true negative – test is negative (indicates [target condition] not present) and patient does not have [target condition]
fn: false negative – test is negative (indicates [target condition] not present) but patient has [target condition]

Figure 2 [colour to be used]: Template for developing test consequence graphic

We suggest only including one flow diagram to summarise the main results to facilitate understanding. Variation in accuracy for each index test from that presented in the flow diagram (for example according to test threshold, quality of studies or differences in characteristics of the population to be tested) can be included in the text.

If multiple summary estimates are available in the review (e.g. more than one index test, different thresholds, different population groups), the following issues need to be considered when deciding which estimate to present as the main results accompanied by a flow diagram. If there are multiple tests, you may choose the test that has the potential to have most impact on clinical practice (e.g. cost, speed of result, invasiveness) or the most accurate test. If multiple summary estimates of accuracy are presented for a single test you may select the estimate of accuracy derived from the threshold that most studies contributed to, the estimate considered most reliable (e.g. restricted to studies at low risk of bias), or the estimate based on the most relevant population (that most likely

to be considered for testing in clinical practice). The most appropriate approach will vary across reviews and will require your judgement and knowledge of the topic area.

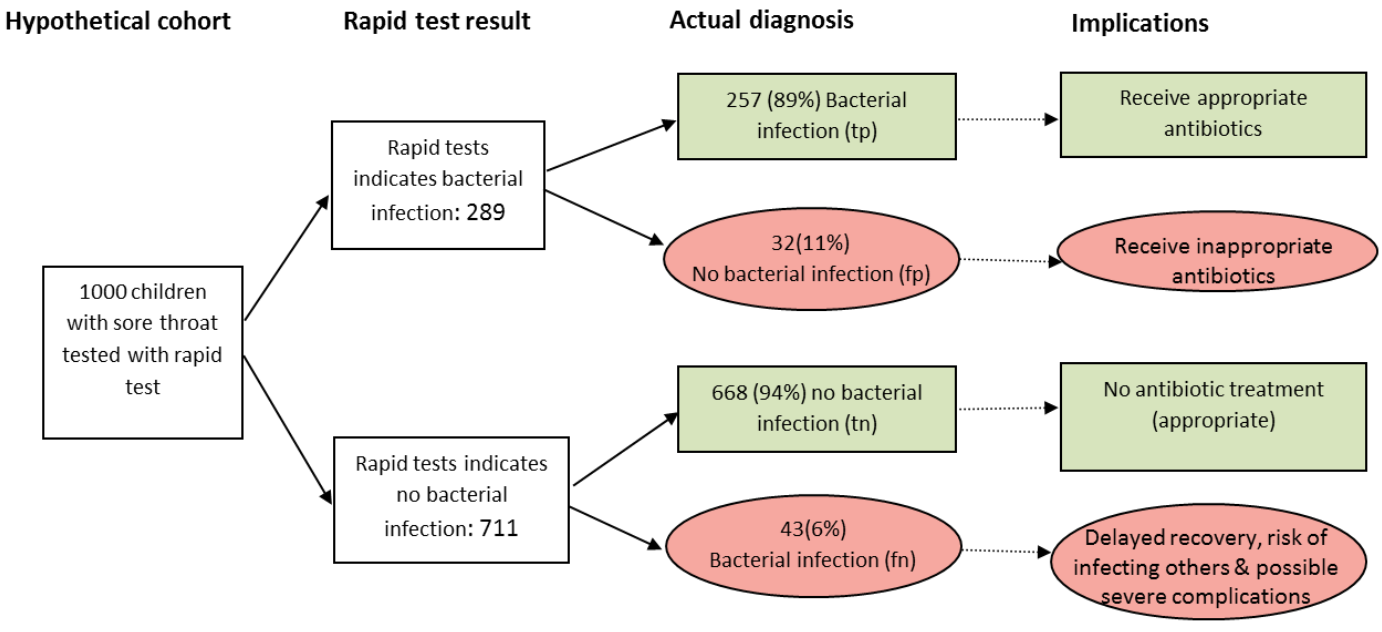
If meta-analysis was not possible or appropriate in a DTA review, then consider whether there are any data that could be presented using the format described above. If this is not possible, a narrative description of results should be presented using natural frequencies.

Example: ***What are the main results of the review?***

The analysis included results from 58 244 children with sore throats. The results of these studies indicate that in theory, if rapid tests were to be used in a group of 1000 children with sore throats, of whom 300 (30%) are actually caused by bacterial infection then:

An estimated 289 would have a rapid test result indicating that their sore throat is caused by a bacterial infection and of these 32(11%) would not have a bacterial infection.

An estimated 711 children would have a rapid test result indicating that their sore throat is not caused by a bacterial infection and of these, 43 (6%) would actually have a bacterial infection.



tp: true positive – test is positive (indicates [target condition]) and patient has [target condition]
fp: false positive – test is positive (indicates [target condition]) but patient does not have [target condition]
tn: true negative – test is negative (indicates [target condition] not present) and patient does not have [target condition]
fn: false negative – test is negative (indicates [target condition] not present) but patient has [target condition]

Figure 3: Test consequence graphic showing results that would be obtained if a hypothetical cohort of 1000 children were tested for Strep A infection using rapid tests

How reliable are the results of the studies in this review?

In this section a summary of the quality of the studies included in the review and the potential impact of bias on estimates of accuracy is presented. The first sentence should describe the reference standard used in the review. A footnote may be used to explain that you are talking about the reference standard for readers familiar with this term, without causing potential confusion by including the term in the text. A comment on whether the reference standard is considered reliable may be helpful. For example:

"In the included studies, the diagnosis of [target condition] was made by assessing all patients with [reference standard].This is likely to have been a reliable method for deciding whether patients really had [target condition]."*

*"*In these studies [reference standard] was the reference standard"*

If there was a potential for risk of bias in the included studies, we suggest using a generic statement and then explaining how bias may have impacted on estimates of test accuracy. We do not recommend going into detail about the type of bias that may have affected the included studies such as verification bias or review bias. For example:

"However, there were some problems with how the studies were conducted. This may result in the [index test] appearing more accurate than it really is, increasing the number of correct [index test] results (green rectangles) in the diagram."

If there is substantial heterogeneity in study results this can be highlighted here. For example:

"The numbers shown in the figure are an average across studies in the review. However as estimates from individual studies included in the review varied we cannot be sure that the [test] will always produce these results."

Lack of precision (wide confidence intervals around summary estimates) and/or small sample size can also be captured in this section. For example:

"Not enough people have been studied to be confident about the results."

Example: ***How reliable are the results of the studies in this review?***

The numbers shown in the figure are averages across all studies in the review. Because result estimates from individual studies varied, we cannot be sure that these rapid tests will always produce the same results. In the included studies, the diagnosis of bacterial infection was confirmed by the most accurate test available: seeing if bacteria could be grown in the laboratory from samples taken from children's throats. Although there were problems with the conduct of some studies, their results did not differ from the more reliable studies.*

Who do the results of this review apply to?

This section should provide a brief summary of the included studies. The mean or median prevalence and range of the target condition across studies should be included. This is particularly important if this is the estimate of the prevalence that was used to populate the figure and to calculate the frequencies for the results section. In addition, a brief summary of pertinent characteristics of included studies should be presented. Information that might be useful includes the countries in which the studies were conducted, details on baseline patient criteria (e.g. symptoms, age, gender, prior tests), the expertise of the person conducting the test.

Example: ***Who do the results of this review apply to?***

The results may not be representative of all children with sore throat being tested in the community. Studies included in the review were carried out in 25 countries with almost half conducted in the USA. Tests produced by many different manufactures were assessed. The average age of children was 7 years. Overall, an average of 29% of children were found to have a bacterial throat infection with this number ranging from 10% to 67% across studies. There was some suggestion that studies in the review included more severely ill children.

What are the implications of this review?

This section provides the conclusions of the review. Start with a brief summary statement regarding whether the results of the review suggest the index test has the potential to be used to detect the

target condition. If the evidence is not sufficient to make a recommendation this should be stated here. E.g. "It is unclear whether [index test] can detect [target condition]".

We suggest next including details on the incidence of false positive and false negative test results and the consequences of these. For example:

"Based on the results of this review, the chance of wrongly diagnosing someone with [target condition] when they do not actually have it appears [comment on frequency e.g. high, low] (XX% of those whose [index test] result suggests they have [target condition]). [comment on consequences]. The chance of missing a diagnosis of [target condition] is [comment on relative frequency e.g. lower, higher] (xx% of those whose [index test] results suggest they do not have [target condition] when they actually have it), [comment on consequences]. These findings should be considered when deciding whether or not to use the [index test] to test for [target condition]"

Details of variation in test accuracy estimates in particular subgroups, for example different patient groups, different test thresholds, different versions of the test, can also be highlighted in this section.

Example: ***What are the implications of this review?***

The studies in this review suggest that rapid tests can detect the most common cause of bacterial infections (Strep A) in children with sore throats, leading to early and appropriate treatment with antibiotics. Both types of rapid tests studied in the review had similar accuracy. The risk of missing a diagnosis of bacterial infection with rapid tests is low (6% of those whose rapid tests suggests they do not have a bacterial infection) suggesting that only a small number of children with a bacterial infection will not receive antibiotics). The risk of wrongly diagnosing a child as having a bacterial infection is slightly higher (11% of those whose rapid test suggests they have a bacterial infection). This may result in some of these children receiving unnecessary antibiotics. The number of children receiving unnecessary antibiotics following a rapid test is still likely to be lower than the number of children who would receive unnecessary antibiotics if the test is not used.

How up-to-date is this review?

State *when* the review authors searched for the included studies, for instance by saying:

"The review authors searched for and used studies published up to [date]."

Example: ***How up-to-date is this review?***

The review authors searched for and used studies published from January 1980 to July 2015.

Discussion

We have introduced a template and guidance on how to write a PLS for a DTA review. Although the guidance was primarily developed for Cochrane review authors we believe that it is equally relevant to other types of DTA review. As far as we are aware, this is the first such guidance available for DTA reviews. We hope that this will help authors of DTA reviews to communicate the results of their review in a more accessible way, leading to improved implementation of results from these reviews. Although PLS primarily target a lay audience, given the complexity of the metrics involved in the evaluation of diagnostic accuracy, we think that the potential audience for a DTA PLS is likely to be much broader than for reviews of interventions and include other end users such as health professionals, journalists and commissioners. We involved all these groups in the development of the guidance to ensure that the final template and guidance would be relevant for them.

Summarising results in plain language is challenging. To be able to write a PLS, reviewers have to really understand the results of their review. Our proposed PLS template encourages reviewers to consider the key messages from their review, what the potential limitations of their review are, and who results are applicable to. The exercise of writing the PLS can also feedback into how the main results section of the review is structured, and can also help inform what information to present in the scientific abstract. Once a draft PLS has been produced it can be helpful to ask a lay member to comment on this. Ideally the review team will include a lay member who will have some understanding of the review topic and can contribute to developing the PLS. They could be asked to comment on the draft PLS or may be able to help in producing the PLS. An alternative would be to share the PLS with an independent lay person who could comment on their understanding and agreement with the key messages of the review.

To conclude, PLS will enable users to better understand the results and conclusions of DTA reviews. This will increase the accessibility of these reviews for various audiences. This paper provides guidance to review authors on how to write these summaries.

References

1. Santesso N, Rader T, Nilsen ES, et al. A summary to communicate evidence from systematic reviews to the public improved understanding and accessibility of information: a randomized controlled trial. *Journal of clinical epidemiology* 2015;68(2):182-90. doi: 10.1016/j.jclinepi.2014.04.009 [published Online First: 2014/07/19]

2. Leeflang MM, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149(12):889-97.

3. Whiting PF, Davenport C, Jameson C, et al. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open* 2015;5(7):e008155. doi: 10.1136/bmjopen-2015-008155

4. Zhelev Z, Garside R, Hyde C. A qualitative study into the difficulties experienced by healthcare decision makers when reading a Cochrane diagnostic test accuracy review. *Systematic reviews* 2013;2:32. doi: 10.1186/2046-4053-2-32 [published Online First: 2013/05/18]

5. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical decision making : an international journal of the Society for Medical Decision Making* 1991;11(2):88-94.

6. Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *Journal of clinical epidemiology* 2013;66(10):1093-104.

7. Whiting P, Davenport C. Understanding test accuracy research: a test consequence graphic. *Diagnostic and Prognostic Research* 2018;2(1):2.

8. Cohen JF, Bertille N, Cohen R, et al. Rapid antigen detection test for group A streptococcus in children with pharyngitis. *Cochrane Database Syst Rev* 2016;7:CD010502. doi: 10.1002/14651858.CD010502.pub2

9. Harrison JK, Fearon P, Noel-Storr AH, et al. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database of Systematic Reviews* 2015(3):CD010772.

10. Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;332(7549):1089-92. doi: 10.1136/bmj.332.7549.1089 [published Online First: 2006/05/06]

11. Davenport C. Systematic reviews and meta-analyses of test accuracy: developing methods that meet practitioners' needs. University of Birmingham, 2012.

12. GRADEPro [Available from: https://gradepro.org/ accessed 6/5/2018 2018.

Author Contributions:

Conception and design: P Whiting, C Davenport, M Leefland, R Mustafa, N Santesso, I de Salis

Analysis and interpretation of the data: PF Whiting, C Davenport, I de Salis, G Gopalakrishna, G Cooney

Drafting of the article: PF Whiting

Critical revision for important intellectual content: All authors

Final approval of the article: All authors

Obtaining of funding: P Whiting, C Davenport, M Leefland, R Mustafa, N Santesso, I de Salis

Collection and assembly of data: PF Whiting, C Davenport, I de Salis, G Gopalakrishna, G Cooney