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

23rd July 2018

Plain language summary template

In this document, we describe how to write a plain language summary for a Diagnostic Test Accuracy (DTA) Review. We suggest sub-headings and provide a description of the content required under each sub-heading.

A

How accurate is the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) test for dementia in hospital?

B

Why is improving dementia diagnosis important?

Not recognising dementia when it is present (a false negative test result) results in lost opportunities for early help including drug therapies or support for patients and carers. An incorrect diagnosis of dementia (a false positive test result) may result in anxiety, stress, wasted resources and unnecessary investigation and treatment.

C

What is the aim of this review?

The aim of this review is to find out how accurate the IQCODE questionnaire is for diagnosing dementia in hospital settings. Researchers included 13 studies to answer this question.

D

What was studied in the review?

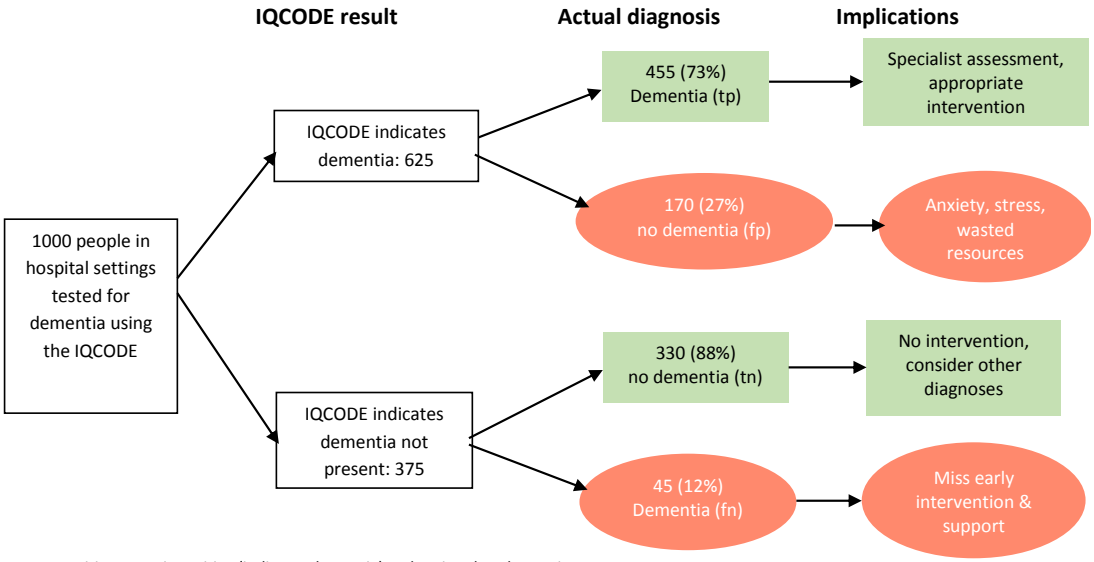
There are two versions of the IQCODE: a "long" version (26 questions) and a "short" version (16 questions). Each question asks if a person's ability to perform certain everyday tasks has changed. Both versions are completed based on information supplied by somebody close to the person being assessed for dementia. Questions are rated on a scale of 1 'has become much better' to 5 'has become much worse'. A diagnosis of dementia is more likely with higher scores.

E

What are the main results of the review?

The review included 13 relevant studies with a total of 2745 participants. The results of these studies indicate that in theory, if the IQCODE were to be used in hospital settings in a group of 1000 people, of whom 500 (50%) have dementia:

- An estimated 625 will have an IQCODE result indicating dementia and of these 170 (27%) will not have dementia
- Of the 375 people with a result indicating that dementia is not present, 45 (12%) will actually have dementia



tp: true positive – test is positive (indicates dementia) and patient has dementia;
fp: false positive – test is positive (indicates dementia) but patient does not have dementia
tn: true negative – test is negative (indicates dementia not present) and patient does not have dementia;
fn: false negative – test is negative (indicates dementia not present) but patient has dementia;

Example 1: This example has been written with the help of the plain language summary template and is based on the following review:
Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. Cochrane Database Syst Rev. 2015 Mar 10;3:CD010772.

F

The IQCODE produces more false positive and false negative results (more people in the red ovals in the diagram) in specialist memory clinics and psychiatry wards than in general hospital clinics and wards. There is no difference in results between long and short versions of the IQCODE or for languages other than English (similar numbers in each box in the diagram above).

G

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

In the included studies, the diagnosis of dementia was made by assessing all patients with an in depth clinical interview*. This is likely to have been a reliable method for deciding whether patients really had dementia. However, there were some problems with how the studies were conducted. This may result in the IQCODE appearing more accurate than it really is, increasing the number of correct IQCODE test results (green rectangles) in the diagram.

H

Who do the results of this review apply to?

Studies included in the review were carried out in Europe, Australia, China, Singapore, and Thailand. Some studies included patients because they had memory problems or other signs of dementia, other studies included general patients admitted to hospital. Average age ranged from 65 to 82 years. The percentage of people with a final diagnosis of dementia was between 11% and 87% across studies (an average of 51%).

I

What are the implications of this review?

The studies included in this review suggest the IQCODE can identify adults over 60 years in hospital who are at risk of dementia and require specialist assessment. The risk of wrongly diagnosing someone with dementia appears high (27% of those whose IQCODE result suggests they have dementia). However, further specialist assessment should distinguish the people wrongly diagnosed from those who actually have dementia. Although the risk of missing a diagnosis of dementia is lower (12% of those whose IQCODE results suggest they do not have dementia), these people may potentially miss early intervention and support. This should be considered when deciding whether to use the IQCODE test to test for dementia. The results also suggest the IQCODE is likely to be less useful in specialist memory clinics and psychiatry wards than general hospital settings. The short version and different language versions are as accurate as the standard English language long version.

How up-to-date is this review?

The review authors searched for and used studies published up to January 2013.

*In these studies clinical interview was the reference standard

Adapted from: Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. Cochrane Database Syst Rev. 2015 Mar 10;3:CD010772.

Instructions for each part

A 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.

B Suggested sub-heading: "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)?

C Suggested sub-heading: "What is the aim of this review?"

The aim of the review should be stated as concisely and simply as possible. People do not always understand that the results of a plain language summary come from a systematic review rather than a single study. Some also wrongly assume that the review authors have carried out the studies themselves. We therefore suggest that you use an introductory sentence such as:

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

D Suggested sub-heading: "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.

E Suggested sub-heading: "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.

Sensitivity and specificity are often reported in diagnostic test accuracy reviews. The tables below can be used to derive natural frequencies from summary estimates of sensitivity and specificity from the systematic review and an estimate of the prevalence of the target condition. The estimate of the prevalence should also be taken from the systematic review, either the mean or median prevalence across studies, unless there is good rationale for taking an alternative estimate.

	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 = 50% (p=0.5), sensitivity = 91% (sens = 0.91), specificity = 66% (spec = 0.66)

	Disease Present	Disease Absent	Total
Test Positive	500*0.91=455	500-(500*0.66)=170	TP+FP=625
Test Negative	500-455=45	500*0.66=330	FN+TN=375
	1000 *0.5=500	1000 – 500=500	1000

There is also a calculator available in RevMan which can provide the information necessary to populate these tables. We also suggest including the proportion of patients with each test result (positive and negative) who have or do not have the target condition. For patients with a true positive (TP) test result this is equivalent to the positive predictive value, for patients with a true negative (TN) test result, this is equivalent to the negative predictive value:

	Disease Present	Disease Absent	Total
Test Positive	455 (455/625= 73%)	170 (170/625=27%)	TP+FP=625
Test Negative	45 (45/375= 12%)	330 (330/375=88%)	FN+TN=375

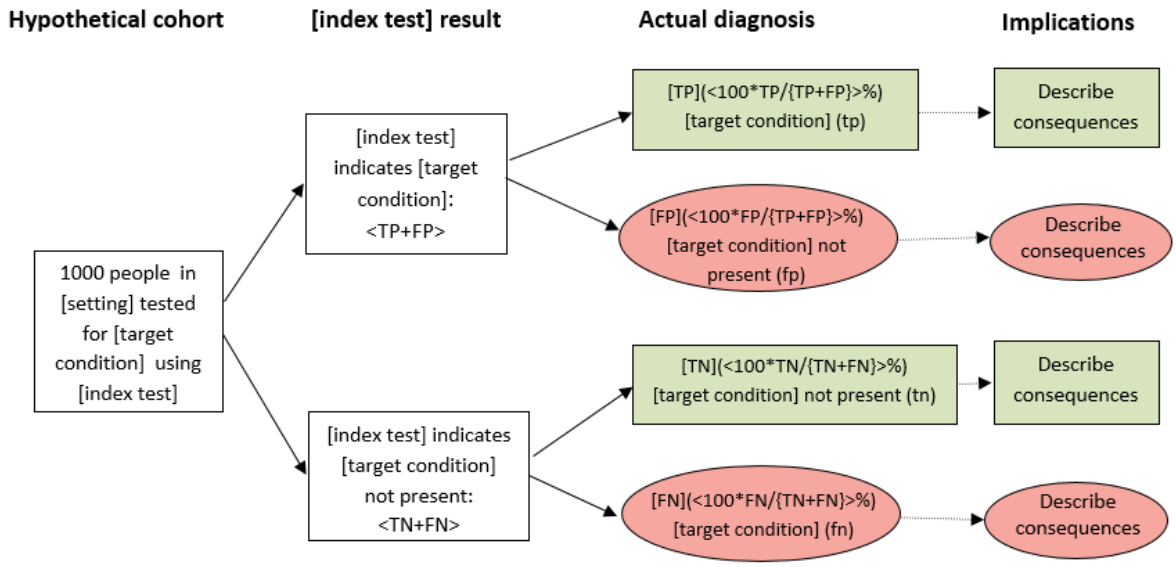
Suggested text for presenting accuracy data:

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# (x%)] have [target condition] then:

- *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 have proposed a flow diagram to summarise these results. This can be populated from the tables above combined with the information on implications that is described under sub-heading B :

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]

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.

F Suggested sub-heading: "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 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.*

G Suggested sub-heading: "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.

H Suggested sub-heading: "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.

I Suggested sub-heading: "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]."*

What are these instructions based on?

These instructions were prepared by Penny Whiting, Clare Davenport and Mariska Leeflang based on the findings of research funded by Cochrane and drawing on the PLS guidance for Cochrane Intervention Reviews. We would like to thank the following for their contributions to this project: Isabel de Salis, Reem Mustafa, Nancy Santesso Gowri Gopalakrishna, Geraldine Cooney, Emily Jesper, and Joanne Thomas.

Additional example: How accurate are rapid swab tests for strep throat in children?

Why is improving the diagnosis of bacterial infection important?

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

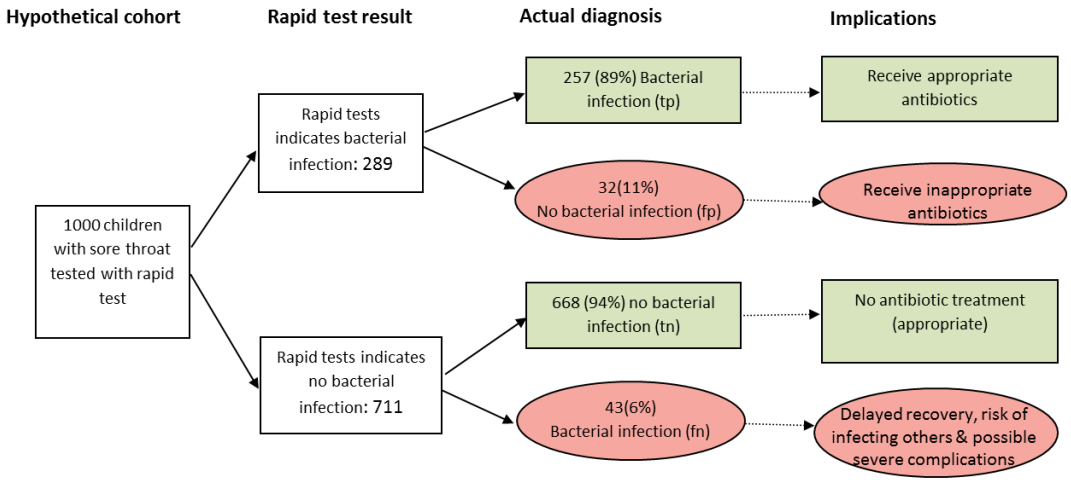
What was studied in the review?

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 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%) 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.
- Both types of rapid test showed similar results.



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]

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?

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?

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?

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

*In these studies, culture was the reference standard

Adapted from: Cohen JF, Bertille N, Cohen R, Chalumeau M. Rapid antigen detection test for group A streptococcus in children with pharyngitis. Cochrane Database of Systematic Reviews 2016, Issue 7. Art. No.: CD010502.