

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How well do health professionals interpret diagnostic information? A systematic review
AUTHORS	Whiting, Penny; Davenport, Clare; Jameson, Catherine; Burke, Margaret; Sterne, Jonathan; Hyde, Chris; Ben-Schlomo, Yoav

VERSION 1 - REVIEW

REVIEWER	George Bergus Carver College of Medicine, University of Iowa, Hawkins Drive, Iowa City, Iowa USA
REVIEW RETURNED	03-May-2015

GENERAL COMMENTS	<p>This is an interesting read. My first impression is "A systematic review about health professionals' interpretation of diagnostic tests, why did I not think of doing this?" Many of us quote studies about the lack of physician understanding about test accuracy. But I am not aware of a prior attempt to systematically review these studies.</p> <p>Introduction Page 6, paragraph 1: We are told that "correct diagnosis is a pre-requisite for appropriate management". This seems obvious but a supporting citation would strengthen this pivotal assumption. I believe Olga Kostopoulou has a study which could be referenced. Page 6, paragraph 2: The authors explain test accuracy. But as they point out, many health professions get lost in these terms. It might make their work more accessible to physicians if the authors provide a 2 x 2 table and define measures of accuracy using the table.</p> <p>Methods Page 7, paragraph 3: The authors report that they contacted experts to identify relevant articles for their review. Can they expand on how they identified these experts? Our group wrote one of the articles included in this review but I do not recall receiving an email from these authors requesting that I identify other relevant articles.</p> <p>Results Page 10, paragraph 1: We are told that "practicing physicians were less able to correctly define sensitivity and specificity compared to medical students and research doctors but exact values were not reported." Without additional information about the methodology of this study it is not possible to make very much sense of this statement. Could the authors either provide this additional information or delete this sentence. Page 10, paragraph 2: The content on self-reported understanding of accuracy measures is very interesting and I would move this content to page 9 and before the text about their actual understanding. This would give better contrast to the difference</p>
-------------------------	---

	<p>between self-perceived understanding of test accuracy and actual understanding.</p> <p>Page 11, paragraph 2: We are told that physicians tended to overestimate post-test probability. This statement does not align the literature on conservatism of probability revision (see Edwards W. Conservatism in Human Information Processing. In Daniel Kahneman, Paul Slovic and Amos Tversky. . Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press; 1982). Can this discrepancy be discussed in the discussion?</p> <p>Page 13, paragraph 2: The authors report on the influence of presentation format on probability revision. While this content is very important I am not convinced it belongs in this manuscript. One concern is that the search strategy might not have been adequately tuned to identify all the relevant articles. For example the study by Cosmides and Tooby is not referenced (Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under un- certainty. Cognition. 1996;58:1-73). Instead I believe it deserves this question needs its own search strategy and manuscript which I hope the authors will undertake in the near future.</p> <p>Discussion</p> <p>Page 15, paragraph 1: The authors tell us that “ the difficulty associated with interpretation of summary test accuracy measures is likely to be a function of their relative complexity.” Could the authors simplify this complicated sentence? I am not sure what this means.</p>
--	--

REVIEWER	Xavier Bossuyt Catholic University Leuven
REVIEW RETURNED	12-May-2015

GENERAL COMMENTS	<p>The authors present a systematic review on the understanding of parameters of diagnostic test performance by health professionals. The manuscript is well written and gives a good overview of the subject.</p> <p>No major comments.</p> <p>Minor comments:</p> <p>On page 12 it is stated: "two studies included an additional scenario in which the likelihood ratio information was provided graphically". This is correct for reference 33. However, in reference 37, the graphical information was related to post-test probability in function of pre-test probability and not likelihood ratio "strictu sensu".</p> <p>Page 13 second paragraph: it could be good to give some explanation of natural frequency, normalised frequency and probability.</p>
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1, George Bergus	
<p>Introduction</p> <p>Page 6, paragraph 1: We are told that “correct diagnosis is a pre-requisite for appropriate management”. This seems obvious but a</p>	<p>Thank you for this suggestion. We have included a reference to the following paper:</p> <p>Kostopoulou O, Oudhoff J, Nath R, et al. Predictors of diagnostic accuracy and safe</p>

supporting citation would strengthen this pivotal assumption. I believe Olga Kostopoulou has a study which could be referenced.	management in difficult diagnostic problems in family medicine. <i>Medical Decision Making</i> . Sep-Oct 2008;28(5):668-680.
Page 6, paragraph 2: The authors explain test accuracy. But as they point out, many health professions get lost in these terms. It might make their work more accessible to physicians if the authors provide a 2 x 2 table and define measures of accuracy using the table.	We have included a box showing this information.
Methods Page 7, paragraph 3: The authors report that they contacted experts to identify relevant articles for their review. Can they expand on how they identified these experts? Our group wrote one of the articles included in this review but I do not recall receiving an email from these authors requesting that I identify other relevant articles.	This comprised presentation of the results of the literature search at a national conference (Methods for Evaluating Tests and Biomarkers: second international symposium, University of Birmingham, July 2010). During the course of the research literature was obtained passively through discussions with experts at national and international conferences and meetings concerned with test evaluation and based on included studies. In addition a search was conducted for authors who had published any empirical research concerning understanding of test accuracy by health professionals or who had published substantively in related areas.
Results Page 10, paragraph 1: We are told that “practicing physicians were less able to correctly define sensitivity and specificity compared to medical students and research doctors but exact values were not reported.” Without additional information about the methodology of this study it is not possible to make very much sense of this statement. Could the authors either provide this additional information or delete this sentence.	We have added additional details to this sentence: “A final study, that involved asking participants to identify definitions based on a 2x2 table, reported that practicing physicians were less able to select correct definitions of sensitivity and specificity compared to medical students and research doctors but exact values were not reported.”
Page 10, paragraph 2: The content on self-reported understanding of accuracy measures is very interesting and I would move this content to page 9 and before the text about their actual understanding. This would give better contrast to the difference between self-perceived understanding of test accuracy and actual understanding.	We agree that this suggested structure makes sense and have moved the section on self-reported understanding to come before actual understanding.
Page 11, paragraph 2: We are told that physicians tended to overestimate post-test probability. This statement does not align the literature on conservatism of probability revision (see Edwards	We have read the recommended chapter and feel that although on initial reading there does appear to be a discrepancy, when considering in more detail we feel that our results are

<p>W. Conservatism in Human Information Processing. In Daniel Kahneman, Paul Slovic and Amos Tversky. . Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press; 1982). Can this discrepancy be discussed in the discussion?</p>	<p>actually consistent with this chapter. We have added the following sentences to the discussion to explain this:</p> <p>“Research in the psychological literature has also shown that individuals are often conservative when asked to estimate probability revisions based on Bayes’ theorem. However, this has been shown only to be the case for information having reasonably high diagnostic value. For information with the least diagnostic value, participants are generally more extreme than Bayes’ theorem.{Edwards, 1982 #123} This is consistent with our findings where most examples presented combinations of low pre-test probabilities of disease or values of sensitivity and specificity that were not sufficiently high for ruling in or ruling out disease.”</p>
<p>Page 13, paragraph 2: The authors report on the influence of presentation format on probability revision. While this content is very important I am not convinced it belongs in this manuscript. One concern is that the search strategy might not have been adequately tuned to identify all the relevant articles. For example the study by Cosmides and Tooby is not referenced (Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under un- certainty. Cognition. 1996;58:1-73). Instead I believe it deserves this question needs its own search strategy and manuscript which I hope the authors will undertake in the near future.</p>	<p>We feel that presentation format is an important component of this manuscript. Our search strategy aimed to identify all relevant studies on this topic, although this is not a straightforward area to search. We have had another look at the Cosmides and Tooby paper . This was not included in the review as it did not fulfil our inclusion criteria in relation to participants “<i>health professionals (e.g. doctors, nurses, physiotherapists, midwives), or student health professionals,</i>” . The participants in the Cosmides and Tooby paper were “students at Stanford University... paid volunteers”.</p>
<p>Discussion Page 15, paragraph 1: The authors tell us that “the difficulty associated with interpretation of summary test accuracy measures is likely to be a function of their relative complexity.” Could the authors simplify this complicated sentence? I am not sure what this means.</p>	<p>We have rephrased this sentence as follows: “Difficulty in interpreting summary test accuracy measures is likely to be related to their complexity.”</p>

Reviewer 2: Xavier Bossuyt	
<p>On page 12 it is stated: "two studies included an additional scenario in which the likelihood ratio information was provided graphically".</p> <p>This is correct for reference 33. However, in reference 37, the graphical information was related to post-test probability in function of pre-test probability and not likelihood ratio "strictu sensu".</p>	<p>We have clarified this by adding details of the types of graphical display as follows:</p> <p>"Two studies included an additional scenario in which the likelihood ratio information was provided graphically – one provided the information as a probability modifying plot, {Vermeersch, 2010 #67} the other as a graphic featuring five circles in a row in which an increasing number of circles were coloured black to correspond with increasing positive likelihood ratios or decreasing negative likelihood ratios.{Puhan, 2005 #21}"</p>
<p>Page 13 second paragraph: it could be good to give some explanation of natural frequency, normalised frequency and probability.</p>	<p>We have included a definition of natural frequency, normalised frequency and probability and an additional reference to box 2 which gives examples. We have removed reference to normalised frequency – this is the same as natural frequency but is a less used term and we feel that including this may be confusing.</p>