Author's response to reviews

Title: Application of GRADE: Making Evidence-Based Recommendations about Diagnostic Tests in Clinical Practice Guidelines

Authors:

Jonathan Hsu (hsuj@mcmaster.ca)
Jan Brozek (brozekj@mcmaster.ca)
Luigi Terraciano (terrycom1957@gmail.com)
Julia Kreis (j.kreis@iqwig.de)
Enrico Compalati (enrico.compalati@gmail.com)
Airton Tetelboom Stein (astein@via-rs.net)
Alessandro Fiocchi (alexfioc@gmail.com)
Holger Schünemann (schuneh@mcmaster.ca)

Version: 2 Date: 10 April 2011

Author's response to reviews:

Dear Prof. Wensing,

Thank you kindly for inviting us to resubmit our manuscript entitled: “Application of GRADE: Making Evidence-Based Recommendations about Diagnostic Tests in Clinical Practice Guidelines”. In the following paragraphs we will provide the point-by-point reply to your and the reviewers’ comments. We believe that we were able to address all comments and are appreciative of the comments that have improved the manuscript.

Editor’s comments

1. Please include a competing interests section at the end of the manuscript, before the reference list. If the authors have no competing interests, please state: “The authors declare that they have no competing interests.”

Reply: We have done that. We had previously included this information under “acknowledgment”.

2. Please include an Authors’ Contributions section at the end of the manuscript, before the reference list. We suggest the following kind of format (please use initials to refer to each author’s contribution).

Reply: We completed this section.

Reviewer 1 comments

We appreciate the reviewer’s suggestions, in particular to expand on the concept of diagnostic testing and have done that.

1. Is the question posed by the authors new and well defined? The report is not
directly answering a research question, so this is not fully applicable. The authors
describe a process and the approach to addressing various challenges faced
when developing clinical practice guidelines on the use of diagnostic tools.

Reply: No response required.

2. Are the methods appropriate and well described, and are sufficient details
provided to replicate the work? Not fully applicable as this is a case-study.

Reply: No response required.

3. Are the data sound and well controlled? Not applicable.

Reply: No response required.

4. Does the manuscript adhere to the relevant standards for reporting and data
deposition? Only partly applicable – to the extent that this applies, the answer is
yes.

Reply: No response required.

5. Are the discussion and conclusions well balanced and adequately supported
by the data?

Yes.

Reply: No response required.

6. Do the title and abstract accurately convey what has been found? Yes.

Reply: No response required.

7. Is the writing acceptable?

Yes, although the authors should expand the text slightly to ease the
understanding of non-expert readers.

Reply: We have done this and report on the details how we have done this in our
responses to the specific comments

Specific comments:

Introduction, para1:

The point about patient-important outcomes is key, but I think the current wording
(3rd sentence) is a bit awkward. I would re-phrase – perhaps use a couple of
sentences, if necessary.

Reply: We have done this. We inserted clarifying sentences.

The added section reads:

For example, recommendations about diagnostic tests should consider the
downstream consequences of such tests. That is, accurate diagnosis is a
prerequisite for successful therapy but an accurate diagnosis should also not be
seen in isolation. Establishing a diagnosis does not provide information about
whether a patient or a group of patients benefits from the diagnosis. Such benefit
should be measured in patient-important outcomes that can include disease
related outcomes (e.g. mortality reduction), psychological consequences of
testing as well as resource utilization outcomes. Recommendations about
diagnostic tests should consider whether these outcomes, when taken together,
achieve net benefit and if this net benefit may be worth the associated resources. In the last sentence, I would add "MAY BE" due to the lack of....

Reply: We have done this.

Introduction para 2:
The claim that health care practitioners “often feel compelled to implement guidelines recommendations” is also awkward to me – for two reasons: 1) I’m not sure if this is true, and 2) The line of reasoning from describing a lack of transparent frameworks to claiming that practitioners feel compelled to use the guidelines doesn’t make perfect sense to me – I fail to see the logical link in the argument (there is a need for transparent frameworks, independent of what practitioners feel – no?)

Reply: We agree that this section benefits from rewording. We intended to express that there is a lack of transparent frameworks. Despite this lack of transparency, guideline developers make recommendations (and little is known about how they do this). Guideline users are likely unaware of the lack of transparency and the lack of (the ability of ) guideline developers to follow transparent frameworks. Being unaware and despite the lack of a framework recommendations are followed by practitioners. We weakened our claim about the practitioners feeling compelled to implement guidelines and revised this section. The revised text reads: Despite the lack of applying transparent frameworks in the development of recommendations about diagnostic tests, it is likely that health care practitioners remain unaware of these limitations and implement guideline recommendations that lack transparency about the assumptions underlying the recommendations, including recommendations about the use of diagnostic tests.

Methods, sub-section “Formulating questions...”, para 3: I would think that a new test could also be introduced simply because it represents a likely improvement relative to an “existing diagnostic pathway” – this is not even mentioned as a possibility (assuming, of course, that the gold-standard test is not the standard test in practice).

Reply: We agree that we failed to mention worse test characteristics or performance and included this in the manuscript.

Methods, sub-section “Formulating questions...”, Para 4: I found the last sentence (“For instance, if the benefits of being correctly....) very hard to understand – I had to read it a dozen times, think for 20 minutes, and write several 2x2 tables before I understood. When I finally understood the sentence, I thought it was clear and simple and felt embarrassed over having struggled for so long with it. I suspect that I am representative for many potential readers who are not readily accustomed to thinking about diagnostic tests – assessing such tests is an intellectual challenge (which is the main reason for publishing this paper in the first place!). I propose that the authors expand this sentence to a few more, in order to make it easier for readers to grasp their point - which I think is key.

Reply: We very much appreciate this comment and the honest description of Dr.
Fretheim’s thought process. We tried to expand in a way that this thought process is clearer. After having worked on this topic for a while, many of us, perhaps all, continue to struggle with the conceptual issues because they simply are “complex”. However, as Dr. Fretheim suggests, this was the prime motivation to describe this and other aspects of this work. Without this expansion, however, we will not achieve our aim of improving understanding. The revised text reads: The main challenge in developing recommendations for diagnostic questions is for panels to understand the implications of the diagnostic test and the quantitative information that diagnostic test accuracy data can provide.[4] GRADE, when making recommendations for diagnosis, provides a structured framework that considers the following outcomes: (1) the patient-important consequences of being classified as true positive (TP), true negative (TN), false positive (FP) or false negative (FN), (2) consequences of inconclusive results, (3) complications of a new test and a reference standard, and (4) resource use (cost). For instance, nearly every test inevitably leads to both correctly classified patients (that can be further separated into TP and TN) and incorrectly classified patients (FP and FN). Correct classification is usually associated with benefits or a reduction in adverse outcomes, while incorrect classification is associated with worse consequences (harms), including failure to treat and potentially reduce burden of disease. A guideline panel needs to evaluate whether the benefits of a correct classification (TP and TN) outweigh the potential harms of an incorrect classification (FP and FN). However, the benefits and harms follow from subsequent action and are determined by probabilities of outcome occurrence and the importance of these outcomes to patients (e.g. mortality, morbidity, symptoms, etc.). If the benefits of being correctly classified by the test (as TP or TN) are sufficiently greater than harms associated with being incorrectly classified (as FP or FN), the guideline panel may be inclined to accept lower accuracy of a diagnostic test when recommending its use.

Methods, sub-section “Test & Treatment…”, para 2:

I would like to learn more about how the exercise that was carried out in order to estimate treatment- and testing-thresholds. This is – at least to me – an innovative approach that deserves more space and detail.

Reply: We also believe that this is innovative and appreciate the opportunity to expand. We asked panel members by email to participate in this exercise. They were invited to complete an “exercise to attempt to estimate the thresholds at which a clinician stops testing for CMA and either starts treatment (CMA very probable) or informs the patient/parents that CMA is not responsible for the symptoms (CMA very improbable).” We provide the full details now in appendix 1 so that others can use this or similar approaches. We inserted the following text in the main manuscript: We asked panel members to estimate treatment and testing thresholds specifying a clinical setting based on history, clinical presentation and results of index tests alone (i.e. without performing a reference test, the OFC). In detail, we applied the following process. Together with the exercise to determine the importance of outcomes, we invited panel members by email to participate in an “exercise to attempt to estimate the thresholds at which a clinician stops testing for CMA and either starts treatment (CMA very probable)
or informs the patient/parents that CMA is not responsible for the symptoms (CMA very improbable) using four different scenarios (appendix 1 includes the detailed exercise).” We informed them that “we acknowledge that these thresholds we asked to estimate are subjective and depend on one’s values and preferences. We also acknowledged that the four scenarios we presented were a simplification of real life situations but that this may be an acceptable trade off between comprehensiveness and simplicity. Following a detailed description of concepts about test and treatment thresholds, contextualization for CMA, provision of probabilities for outcomes and cost estimates, we asked participants to determine their test and treatment thresholds for four scenarios that were described in detail (see appendix 1). We utilized the results of this survey to explore where the thresholds for test recommendations are located along the probabilities of 0 to 100%.

Results.

I feel that the first para is more naturally placed in the Methods-section. I propose adding one sentence about how the various pre-test probabilities were selected (I assume more or less out of the air, and that’s OK – but should be reported, still).

Reply: We followed this suggestion and moved this paragraph although we did not select them out of the air. We provide the details for how the estimates were obtained. They were actually based on the data we obtained from the systematic reviews which is now included in the manuscript.

Box 1: Was difficult to read in the version I received, so I have not reviewed that properly. It’s an important element, though, as I expect guideline developer may use what’s in the box as a practical guide. I therefore urge the authors to carefully review the Box to ensure that it includes all important elements in the process – in my mind this includes the need to specify pre-test probabilities, treatment/test thresholds etc.

Reply: We apologize for the poor readability of this box. We have expanded it and it should read fine now. We have followed the suggestions above (actually these steps were largely included in the box).

Reviewer 2

Major Compulsory Revisions

This is a worthful and nicely reported exercise. The authors report on using the GRADE format in one specific situation. There is an overlap between the GRADE authors and the authors of this paper. This is not wrong nor problematic, but the reader should know this as the paper authors can not be independent when judging the GRADE. I would suggest to clearly mention this on the Introduction or Methods section.

Reply: Thank you for these comments. We have added our involvement in the GRADE working group in the Methods section. We have also made this clearer in the competing interest section of the manuscript.

The revised text (at the beginning of the methods section) reads:
General methods
We conducted a case study based on written records, meeting minutes and critical analysis of the process used to develop the WAO CMA guidelines. Three of the contributors to this article (HJS, JLB and JK) are members of the GRADE working group and have, to a varying degree, contributed to the development of the GRADE approach.

Minor Essential Revisions
Abstract & discussion section: After this exercise, do you conclude that the GRADE is good and can be proposed to be used or not?
Reply: Given our involvement in the development of GRADE, we had refrained from drawing these conclusions but have now inserted the following text: “This case study provides useful guidance for guideline developers and clinicians about what they ought to demand from clinical practice guidelines to facilitate implementation and strengthen confidence in recommendations about diagnostic tests.”

Methods / conflict of interest: 3th line: "recuse"?? Do you mean excuse?
Reply: We used the term recuse deliberately.

Methods / Formulating questions / 5th §: How did you take the asymmetry of diagnostic tests into consideration? In real life FP can either be much worse or less important than FN, dependent on the specific situation.(you correctly did this in the next paragraph).
Reply: We agree with the asymmetry. This was taken into account by explicitly asking to rate the importance of FP and FN as well as describing the consequences related to the FP and FN (as is then provided in the subsequent paragraphs and in the revised text in the methods section – please see response to reviewer 1 regarding paragraph 4).

Results, 3th § / second half: To this reviewer the second half of this § is difficult reading. Please clarify.
Reply: We agree that this paragraph was unclear (now paragraph 2 in the Results section). The revised section reads: The test and treatment threshold work up yielded variable results from guideline panel. Some panel members indicated that they were not willing to accept any residual uncertainty about the presence of CMA. This aversion to any uncertainty is evidenced by providing high treatment thresholds when only the index tests were used. These high treatment thresholds identified a type of clinician who would always perform reference test (OFC). As a consequence, recommendations were made expressing that for settings where OFC would always be performed, index tests would be redundant given their limited accuracy and should not be used.

We hope that we addressed the comments appropriately. Thank you for your consideration.

With kind regards,
Holger Schünemann, M.D., Ph.D., M.Sc., FRCPC
Chair, Department of Clinical Epidemiology & Biostatistics
Michael Gent Chair in Healthcare Research
Professor, Dep'ts. of Clinical Epidemiology & Biostatistics and Medicine
Office contact information:
Department of Clinical Epidemiology & Biostatistics
McMaster University Health Sciences Centre