Title: Overview of the effectiveness of reminders in improving professional behaviour

Authors:

Amy Cheung (dramy.cheung@gmail.com)
Alain Mayhew (almayhew@ohri.ca)
michelle Weir (Michelle.c.weir@gmail.com)
nicole kozloff (nkozloff@gmail.com)
kaitlyn brown (kaitlynbrown901@hotmail.com)
jeremy grimshaw (jgrimshaw@ohri.ca)

Version: 2 Date: 23 May 2012

Author's response to reviews:

Dr. D. Moher
Editor in Chief
Systematic Reviews
May 23, 2012

Dear Dr. Moher,

Re: Revisions on manuscript “Overview of systematic reviews of the effectiveness of reminders in improving healthcare professional behaviour”

Please find below our responses to the reviewers’ comments for our manuscript. Thank you again for your consideration of this manuscript.

Sincerely,

Amy Cheung, MD
Comment 1:
I thought this was interesting and worth doing and my main comments relate to
the need for more clarity about the potential biases introduced by doing a review
of reviews like this. Also, because we are at one additional remove from the
primary studies, some of the basic contextual information is missing, like what
the settings for the primary studies is, and what the study designs are of the
primary studies that are included in the reviews. Presumably this basic
descriptive information about the original studies is available in the reviews
themselves, but by the time we get to reviews of reviews, it is lost and “the
evidence of effectiveness” starts to look quite abstract. So in short it would be
useful to know a little more about the original studies.

Response:
Thank you for this suggestion. We have highlighted further in the Limitations
section that discuss this issue. In this overview, we are unable to provide further
details about the primary studies because there is very little data in the reviews
themselves regarding the primary studies. Of course, this doesn’t limit readers
from going back to original studies.

Comment 2:
The overlap issue is a major one, and the combination of this and votecounting
has the potential to introduce significant bias. For the reader however this is
difficult to judge, as we are not given much information on the actual extent of the
overlaps. What exactly is the overlap between the high-quality reviews in terms
of their primary studies? Is doublecounting really a problem? At worst, exactly the
same studies appear in all the reviews; at best, there is no overlap. Presumably
the truth lies somewhere in between... but we are not told. For example, there is
a risk that “popular” interventions are more likely to be subject to repeated
evaluations, and appear in several reviews, so the evidence for those (based on
votecounting) will appear to be stronger than it really is. I felt that the paper
needs to this in more detail so as to give the reader a better sense of what if any
bias is being introduced.

Response:
Thank you for raising this issue for further discussion. We have now detailed the
extent of the doubling counting in the results and its possible impact in the
discussion.

Reviewer: Duncan Chambers
Major compulsory revisions

Comment 3:
The authors should explain and justify why they considered reviews with an AMSTAR score >5 to be high quality. As far as I can tell the AMSTAR papers themselves don’t give any guidance on this but Rx for Change classifies scores of 4 to 7 as medium quality and only those with 8 or more as high quality.

Response:
Thank you for the comment regarding the categorization of reviews based on the AMSTAR score. We have changed the wording in the manuscript to reflect the correct categorization of reviews based on their AMSTAR score into high, medium and low quality reviews. We could not limit our overview to only high quality SR’s because there was only one that scored 8 or above.

Comment 4:
Linked to this, the summary score by itself isn’t very informative. It would be helpful to have the answers to the individual questions for each review. This could be added to the additional table; a shorthand version would be fine.

Response:
We have attached a summary table of individual AMSTAR items (Table 1) and frequency of "yes" scores on each item for the 35 reviews. We have also revised the text which refers to these individual items.

Comment 5:
Vote-counting as a method of synthesis clearly has major limitations, i.e. it takes no account of sample size, study quality or the magnitude of effect. I think the authors need to acknowledge these limitations and justify the approach they have taken. The Cochrane handbook chapter cited by the authors recommends attempting to assess the strength of the evidence for each major outcome. Would it not have been possible to do this for the primary outcome, perhaps using a GRADE-type approach?

Response:
Thank you for the suggestion for using GRADE. Unfortunately, none of the reviews used a GRADE type approach. Furthermore, the methodology for using GRADE in overviews of reviews is not worked through yet. The justification of vote counting has been expanded in the Methods section.

Comment 6:
Conclusions, first sentence: ‘reminder systems are more likely to be successful if they are patient specific and meet the specific needs of the clinical setting...’. This rather seems to come out of nowhere given the limitations of the reviews focused on settings and patient groups. Can the authors justify the statement more specifically using the evidence presented in the overview?

Response:
Thanks for pointing out this error. The text has been changed accordingly.

Minor essential revisions

Comment 7:
The authors should clarify their reporting of the search. The abstract states that MEDLINE, EMBASE, DARE and the Cochrane Library were searched but from the Methods it appears that these databases were searched indirectly via Rx for Change. In practice it probably wouldn’t make much difference but the abstract could give the impression that MEDLINE etc. were searched specifically for this overview. The statement in the second paragraph of the discussion that the overview was informed by ‘a comprehensive search strategy...’ rather adds to the confusion.

Response:
Thank you for the suggestion. The text in the abstract and the manuscript have been changed to reflect more accurately the purpose of the original search.

Comment 8:
Report whether a protocol was drawn up in advance for this overview.

Response:
No written protocol was drawn up in advance and this statement was added in the text.

Comment 9:
Figure 2, labelling of years needs improving.

Response:
The figure has been relabeled.

Comment 10:
Figure 3, it would be more informative to present actual numbers (as in Fig. 2) rather than percentages.

Response:
The figure has been changed to include numbers and not percentages.

Discretionary revisions

Comment 11:
Although the overview focused on process outcomes, it would be interesting to know which of the reviews reported health/patient outcomes and how strong the evidence is that reminders can improve these outcomes.

Response:
We agree that other outcomes are important. However, this review is focused on outcomes related to professional behaviours and to include other outcomes...
would entail a brand new search which is beyond our scope.

Comment 12:
Last sentence under ‘reviews of specific settings’: perhaps remove or rephrase
the reference to number of studies as this is not necessarily related to quality of
the review?
Response:
The references has been moved to earlier part of the sentence.

Comment 13:
Under ‘reviews of specific behaviours’, it would be helpful to have some
interpretation of the numerical results from Durieux et al. Was the effect clinically
significant?
Response:
The goal of this overview is to evaluate the evidence for the use of reminders to
change professional behaviour. Therefore, we have presented the data from all
the reviews found and have allowed the readers to determine whether this is
clinically meaningful – both the results of individual reviews and also the
overview.

Comment 14:
Under ‘reviews of specific patient populations, the authors could comment on the
apparent discrepancy between the results of the overall vote count and the
findings of the higher quality reviews.
Response:
The different findings may be related to quality but also to two other factors. First,
when we focus on specific patients, we are limiting the number of studies
included. With the smaller sample of studies, we are less able to determine
whether differences in the results are related to true differences or simply to
chance. Second, there may be true differences due to patient characteristics.
However, due to the small number of reviews, we were also not able to make this
determination.

Comment 15:
The discussion (last paragraph) could reflect on why more narrowly focused
reviews are done, i.e. they are more likely to be helpful to decision-makers.
Response:
The research literature is mixed as to whether policy makers like broad or narrow
based reviews. In fact, one of our co-authors have a manuscript in press
addressing this specific issue and found poor justification about lumping (broad
based) or splitting (narrow based) in reviews of reminders (Weir et al., in press).