Author's response to reviews

Title: Incident reviews in UK maternity units: a systematic appraisal of the quality of local guidelines

Authors:

Anjali Shah (Anjali.shah@npeu.ox.ac.uk)
Olaa Mohamed-Ahmed (Olaa.mohamedahmed@npeu.ox.ac.uk)
Philippe Peirsegaele (Philippe.peirsegaele@npeu.ox.ac.uk)
Charlotte McClymont (Charlotte.mcclymont@npeu.ox.ac.uk)
Marian Knight (Marian.knight@npeu.ox.ac.uk)

Version: 3
Date: 12 January 2015

Author's response to reviews: see over
Dear Professor O’Donovan

Re: Incident reviews in UK maternity units: a systematic appraisal of the quality of local guidelines

Thank you for inviting us to resubmit this manuscript after careful consideration, and for having obtained comments from knowledgeable reviewers.

We have answered the reviewers’ queries and made changes to the manuscript as appropriate.

We hope that given the improvements made to the manuscript, you will now find it suitable for publication. We look forward to hearing from you soon.

Yours sincerely

Dr Anjali Shah
(on behalf of co-authors Dr Olaa Mohamed-Ahmed, Philippe Peirsegaele, Charlotte McClymont and Professor Marian Knight)
Response to reviewers

Reviewer 1: Annette Bernloehr

Overall, Shah and colleagues have submitted a very interesting and important paper on the systematic appraisal of the quality of local guidelines for incident reviews in UK maternity units.
1) The research question is clear, interesting, and well defined.
2) The chosen methods are generally appropriate and well described.

*We thank the reviewer for the positive comments.*

Surveying all consultant-led maternity units in the UK is ambitious, but adds a lot to the importance of findings. However, information is provided on how often the individual units were asked for their guidelines, but information is missing on when the survey was conducted (year), and how the contact took place (telephone, e-mail, surface mail, other).

*<A sentence in the first paragraph of the Methods (lines 109-111) has been modified to include the requested information. It now states: ‘All 211 consultant-led maternity units in the UK were contacted up to three times, by postal mail, e-mail and telephone, and asked to supply a copy of their maternity risk management strategy and/or their incident review procedure during 2012.’>*

It is good that an appraisal with the AGREE II instrument was supplemented by analysing the methods used for reviewing incidents, the people involved in such a review and the methods for disseminating the outcomes of the reviews. A sole appraisal of the guidelines without looking at their actual content would not have sufficed. However, it would have been advantageous to address some of the limitations of the AGREE II instrument more explicitly, i.e. that there is no proven link between the quality of a guideline as determined by the AGREE II appraisal, and the quality of the actual content of a guideline. Here, some explanatory words could help the readers of the publication, but especially other researchers planning to use the instrument in the future.

*<We have modified the Methods (line 160) and the Discussion (lines 289-296) to include this point as requested. The Discussion now states: ‘The AGREE II instrument was supplemented by analysing methods used for reviewing incidents, the people to be involved in a review and the methods for disseminating outcomes, because there is no proven link between the quality of a guideline as determined by the AGREE II appraisal, and the quality of the content of a guideline.’>*

Also good is that the individual items of the AGREE II instrument were adapted to the guidelines under review through a discussion of the two researchers.

With regard to the chosen reviewers of the guidelines, several issues arise:

a) Although guidelines can be appraised with the AGREE II by persons without prior knowledge of the clinical specialty of the guideline, it has been found that some knowledge of the specialty leads to different judgements (Bernloehr 2007). It might have therefore been better to introduce an obstetrician or a midwife into the team of appraisers, although it might not be this important when analysing guidelines on incident reviews.
Prior to evaluating any guidelines, the appraisers discussed each question within the AGREE II criteria, and then met with a midwife to resolve a small number of outstanding issues that needed clarification. This point has been added to the Methods (line 153-154).

b) The AGREE collaboration usually recommends four independent appraisers. It would have been suitable to explain, why only two appraisers reviewed the guidelines, and why this was regarded as acceptable.

The AGREE II User’s Manual (May 2009) states ‘We recommend that each guideline is assessed by at least 2 appraisers and preferably 4 as this will increase the reliability of the instrument. Reliability tests of the instrument are on-going.’

Two appraisers reviewed the guidelines and were found to be largely in agreement on the scores for each AGREE II item. A third appraiser (a midwife) was available to review guidelines, but was not needed. Acknowledgment of this issue has been added to the Discussion (lines 278-281): ‘Two appraisers reviewed the guidelines and gave scores that were largely in agreement for each AGREE II item. Having additional appraisers may have increased the reliability of the instrument, but could be unnecessary given the similarity between the independent scores.’

c) Moreover, the procedure of two appraisers discussing and revising their ratings when major differences in ratings were observed is questionable. Such a procedure is known e.g. from systematic reviews, but usually discouraged for appraisals with the AGREE II instrument. However, a person experienced with AGREE appraisals will understand why you introduced this procedure, and what the advantages are. Persons not experienced with such appraisals might not notice this deviation from the recommended practice, but would benefit from some more explanation and a deeper understanding. AGREE II is easy to apply and widely used, but needs careful consideration and explanation when used.

Explanation for these additional procedures has been given in the Discussion (lines 282-286): ‘It is not standard practice to discuss the questions or review scores, but these measures were useful given that neither appraiser had used the AGREE instrument before, some questions could have been interpreted as being repetitive and there were some errors and discrepancies in interpretation.’

d) One section under the headline ‘Quality appraisal’ is not clear (line 138ff). Was an overall mark for a guideline’s quality calculated? Here, the explanation of what was actually done is not understandable. However, in the results chapter, average AGREE scores are reported (line 211ff). Unfortunately, calculating overall scores for guidelines is straightforwardly discouraged by the AGREE collaboration. Deviating from what is recommended by the developers of an instrument clearly needs to be addressed and justified.

An overall mark for a guideline’s quality was calculated by producing the mean average score from the 23 scores for each AGREE criteria. Although this approach is not recommended by the AGREE collaborators, we found that the subjective approach recommended by AGREE did not capture the variability in quality between the guidelines (a score of between 1 and 7, as well as then recommending, recommending with modifications or not recommending a guideline). Thus we used an objective, quantitative measure (mean average score) to categorise guidelines into one of four
groups: poor quality, average quality, good quality and high quality. There was a higher concurrence of scoring between the two reviewers using the objective approach.

The following sentence has been added to the Discussion (lines 292-296): ‘An overall mean average score was calculated for each guideline and then guidelines were categorised as poor, average, good or high quality to provide an objective, quantitative measure of the quality of a guideline. Current guidance suggests a more subjective approach that we found did not capture variability in quality between guidelines.’

3) The presented data are sound and relevant, although the calculation and presentation of ‘average AGREE scores’ for guidelines is questionable (see comment 2d).

Please refer to response for question 2d.

4) Figures represent the findings of the study and do not show any evidence of manipulation.

5) Relevant standards for reporting data are adhered to.

6) Discussion and conclusions are well backed up by the results of the study, balanced, and discussed in the light of previously published evidence. The discussion on the domain ‘rigour of development’ is good. However, it could be generally questioned whether the AGREE II instrument is suitable for appraising adaptations of (national) guidelines.

We agree with this point and have added it to the Discussion (lines 316-318), which now reads ‘Failure to include a score for the use of national guidance positively in the ‘rigour of development’ domain may be regarded as a potential limitation of the AGREE II instrument. It could be generally questioned whether the AGREE II instrument is suitable for appraising adaptations of (national) guidelines.’

7) Limitations of the work are addressed, but the limitations of the chosen methods and how they were applied could be described better (see comments 2a-d).

Please see response to comments 2a-d.

8) The work of others is clearly acknowledged and integrated into the work.

9) Title and abstract are fine.

10) Academic writing is of high standard. One section under the headline ‘Quality appraisal’ is not clear (line 138ff), see comment 2d.

Please see response to comments 2d.

Discretionary revisions:
It is suggested to add the key words ‘AGREE’ and ‘guideline(s)’.
These suggestions have been included in the key words.

It is recommended to add the missing explanations on the methods, as detailed under 2a-c. Please see response to comments 2a-c.
Minor essential revisions:
Add the missing information on when the survey was conducted (year), and how the contact took place (telephone, e-mail, surface mail, other).
<The missing information has been added as described above.>

References 5 and 7 are identical. Delete one of them and re-number the others, please.
<This amendment has been made.>

Major compulsory revisions:
Comment 2d needs consideration.
>Please see response to comments 2d above

References:

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests: I declare that I have no competing interests

Reviewer 2: Christiane Schwarz

Discretionary Revisions
1. Is the question posed by the authors well defined?
The context of the work is clear, and the question is logical. If lessons can be learned from incidents, the cases need systematic analysis and follow-up. This requires clear guidance of high quality. The authors conduct a systematic search within the UK, and appraisal with an appropriate instrument. The clear definition of the kind of incidents the guidelines deal with becomes clear on second sight only for individuals who do not work in the UK. In some countries, there may be a different meaning of „critical incident“ („near miss“), adverse incident/ event („patient harmed“) and „sentinel event“ (i.e. maternal death). Obviously, all of the above are defined as „incidents“ in this article (according to NPSAFramework). Maybe the authors could add a few words on this definition.
<As discussed in the background section, individual units have different guidance on what constitutes a critical incident. We have further clarified by adding the following sentence to the ‘Background’ section (lines 94-96): ‘The definition of a maternity ‘incident’ that should trigger a review thus varies across the UK.’ We hope this will provide clarity on the context in the UK.

2. Are the methods appropriate and well described?
The AGREE II instrument is a validated tool for the authors’ purpose, to appraise guidelines on a specific topic. AGREE II includes good guidance on how to systematically appraise each domain of a guideline, which makes the results transparent and reproducible by peers.
3. Are the data sound?
The data appear sound and trustworthy. The fact that the authors benchmarked the trusts that were examined in their study with those that did not reply adds to the quality of the results.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
Yes.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?
The manuscript mainly follows the standards the appraisal process that the AGREE consortium suggests, with the exception that “average scores” are reported (lines 211-217) while AGREE suggests to rate the domains independently.

*Please see our response to this point (above), which was also made by the first reviewer.

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

7. Are limitations of the work clearly stated?
The authors refer to the fact that not all trusts replied to their request.

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
The authors did acknowledge the existing body of evidence. The key publication NPSA Framework for Reporting and Learning from Serious Incidents Requiring Investigation is quoted as 2010 version, although a 2013 update version is available.

*The updated version of this document has been cited in reference 16.

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

10. Is the writing acceptable?
The writing is perfect. Tables and figures are appropriate.

11. Other:
Some UK-specific abbreviations need clarification: what does a CNST level of 1 mean – does that mean good, or bad quality, or something completely different?

*The following sentence has been added to the Methods (lines 115-119): ‘A CNST level of 1 indicates that a unit has a process for managing risks that has been documented, a CNST level of 2 indicates that a unit has been assessed as following the process and a CNST level of 3 indicates that a unit monitors the process for managing risk, identifies deficiencies and draws up action plans to reduce risks.’

Bibliography:
- 5. and 7. appears to be the same publication?
*This duplicate has been removed.
- 8. the link is broken

<We successfully tried the link, which takes you to the system where an incident can be reported. It is not a specific document.>

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests: I declare that I have no competing interests