Author’s response to reviews

Title: Clinical drug trials in general practice: How well are external validity issues reported?

Authors:
Anja Maria Brænd (ambraend@medisin.uio.no)
Jørund Straand (jorund.straand@medisin.uio.no)
Atle Klovning (atle.klovning@medisin.uio.no)

Version: 1 Date: 30 Aug 2017

Author’s response to reviews:

FAMP-D-17-00187

Clinical drug trials in general practice: How well are external validity issues reported?

Anja Maria Brænd, MD; Jørund Straand, MD, PhD; Atle Klovning, MD, PhD BMC Family Practice

Dear Editor

Thank you for the opportunity to submit a revised version of our paper. Both reviewers have contributed with valuable suggestions and remarks, and we believe the additional work done in the review process has improved the manuscript, especially regarding the referenced literature and the statistical analyses. In the following we give a point-by-point response to the reviewers’ comments. We hope you will consider the revised paper worthy of publication in BMC Family Practice.

Best regards,

Anja Brænd on behalf of all authors

Reviewers’ Comments to Author:

Merrick Zwarenstein (Reviewer 1):

Dear Authors, I think this is a well conducted review of a large number of relevant trials, and would be immediately publishable if you somewhat improve your literature review and include other existing works on this same topic in your background and discussion.

Specifically, your assertion that there is no existing standard for reporting trials with reference to their external validity or generalizability is not correct. I think the introduction and discussion
sections of your paper should be rewritten to take critical account of the already published standards for this question, and your empirical results can be used to argue for or against the comprehensiveness of the existing published standards for external validity.

On the Consort website you will find a Consort reporting standards paper specifically aimed at reporting pragmatic trials, that is, trials aiming at wide applicability and real world use of results.

The published reference is:


* Thank you for this remark and the reference, which we have included in the introduction section. Our cohort of trials includes trials with a pragmatic approach, but also trials of a more explanatory character, even though the general practice setting first and foremost invites for pragmatic trials. We agree that the discussion of the existing literature was somewhat insufficient, and we have revised the paragraph in the introduction section regarding reporting of external validity issues. Recent publications have also been included in the referenced literature, and we also mention the PRECIS tool in this section. The references in the discussion section have also been revised. However, we believe that a comprehensive assessment of all existing published standards for external validity is beyond the scope of this paper, which is extensive enough as it is.

On the same note: you reference another paper in which I was involved, but don't really grapple with its main point, which is explicitly about designing trials for external validity. This paper, your reference number 50, itself contains a series of 9 recommendations for items that should be considered during design of a randomized trial; these items could well be used as a basis for your discussion on what should be reported for trials with an intention to focus on real world usefulness and external validity.

* We agree that a presentation of the PRECIS tool is relevant, and we now introduce the tool in the background section.

Anne-Kirstine Dyrvig (Reviewer 2):
The manuscript concerns an aspect of utmost importance and relevance. I find that there is sufficient coherence between the themes given in the title, abstract, manuscript, and conclusions.

My main concerns and the reason that I selected the option "major revisions" is that the statistical analyses need to be reconsidered and that some central aspects in terms of organising the paper and restating the aims is necessary. Otherwise I truly enjoyed reading the paper and find that the Work is valuable and a great contribution to the field of external validity.
* We wish to thank the reviewer for this acknowledgement.

My comments are structured in chronological order with reference to the manuscript section below.

Introduction:

- The aim is clearly stated, although it would be valuable with a definition of "sufficient information" or perhaps choosing another term. I would recommend that a section is added with clear specified objectives of the paper, as the aim does not cover all the topics/objectives that are addressed in the manuscript.

* Thank you for this remark. We have now included specified objectives as suggested.

- Also, please consider if your ambitions with the manuscript are too extensive for one article, or if it could be better presented in two manuscripts. For instance, the "case study" that is used partly for illustrative purposes could perhaps provide sufficient information for a separate article.

* We acknowledge that the scope of the paper might be too extensive, and we have considered removing the case study. However, we believe the case study gives a good illustration of the characteristics of included patients for the largest group of trials compared to other published data.

Methods:

- Study cohort is clearly described. It is not clear, however, why the different cut-offs were made (period 1998-2007 and papers with publication years 200-2015). It would be nice to know why these periods were selected and why the delimitation was made on period rather than publication or on whether studies were finished or any other characteristic.

* The 10-year period 1998-2007 was chosen partly because there was a conversion to electronic registration in the Norwegian Medicines Agency archive after 2007, which represented an obstacle regarding permission to electronic access etc. The time period was also chosen to allow sufficient time for results to be published after the trials were ended. All published papers that were identified from the cohort trials were included, comprising the 134 papers. These papers were published in the period 2000-2015. The last search for publications was conducted in December 2015, and we have described the publication output from the cohort of trials in a previous paper (reference 25), but papers were not excluded otherwise. It is possible that further papers may have been published after the last search for papers was conducted in December 2015, but we consider it unlikely that this would affect our results. We have added a sentence in the methods section to clarify this.
- Data extraction: Why were only half of the trials assessed by two authors. And how were they selected (it is mentioned in the discussion that the selection was random)

* For reasons of feasibility not all papers were assessed by two persons. To safeguard the data extraction we chose to double code a random half of the papers. For these papers, data were extracted independently by another author, and Kappa statistics for agreement were calculated and found to be good. We used an internet based random number generator (www.random.org) to select the random sample based on the trial identification number. We have included a clarifying statement regarding this in the methods section.

- Data extraction: You mention that data are extracted on setting, patient selection, primary outcome, and eligibility criteria. While those are reasonable choices, they are not properly introduced and thus it seems they were not actively chosen. If they were selected, please mention if they were selected on the basis of availability, relevance, or...?

* Thank you for this comment. We have rewritten this section, hopefully clarifying the rationale for our selection.

- Data extraction: Here the case-study is mentioned. It should be introduced earlier. And if the current amount of attention spent on the case-study is to be maintained, the case study deserves an earlier introduction, and a mentioning in the aims/objectives.

* We agree, and we have now included the case study in the objectives.

- statistical analyses: The section is fine, although it is unclear why the choice was made to test for before/after 2007. I think these tests that were conducted for a number of the variables are arbitrary. Either the variables should be tested for trends over time, or analyses should be kept to descriptive statistics. Chi-squared tests for arbitrary 2X2-tables seem like an attempt to introduce statistics, where statistics may be redundant as the trends can be spotted through the very nice Graphs presented in the appendix.

* We agree that testing for differences before/after the median year has limitations and may seem to some extent arbitrary. It was chosen to get a numerical expression of what is shown in the graphs, as you point out. We have reconsidered these analyses and have performed Chi square tests for trend, according to Altman: Practical statistics for medical research (p. 261). The results corresponded well with the results of the previously performed 2x2-tests. We have removed the 2x2 results from the text and table 1 and replaced with the Chi square tests for trend. We have also specified that exploring development over time was an objective, but as suggested, we have given less attention to this in the results and discussion sections.

Results

- Reporting of trial setting: Much attention is given to the presence or absence of statistical significant differences. But as it was not mentioned as an aim to investigate the
Development over time (or, more precisely, the difference between before and after 2007), it seems quite surprising and not necessary.

* We agree that the statistical significant differences might have been overemphasized (see above). We have now included the investigation of development over time as an objective.

- I would suggest to simply report the items that were investigated and drop the focus on before/after 2007.

* See above.

- Reporting of patient selection: You state in the final sentence: "The proportion reporting eligibility criteria improved over time", but is that supported with an analysis or is it mainly based on visual interpretation of the figure?

* See above.

Discussion

- The section is generally difficult to understand, as it is not clear when authors are discussing their own results or when they are referring to the results of others. Revision of the language is recommended, as the points made in the section are valid and relevant.

* We have restructured the discussion section and revised the language. Hopefully, this has clarified when we discuss our own findings and when we discuss existing literature. Since language revision was recommended, the manuscript has also been reviewed by an English language editing service.

Conclusions

- The sentence "we here observed improved reporting over time". I don't think this is true on the basis of the analyses. You observed an improvement after as compared to before 2017. Not necessarily a trend over time.

* We believe the Chi square test for trend supported that we observed an actual improvement in this reporting over time, and therefore we have kept this statement, slightly rephrased.