Author’s response to reviews

Title: Insufficient uptake of systematic search methods in oncological clinical practice guideline: a systematic review

Authors:

Chiara Trevisiol (chiara.trevisiol@aulss3.veneto.it)
Michela Cinquini (michela.cinquini@marionegri.it)
Aline Fabricio (aline.fabricio@aulss3.veneto.it)
Massimo Gion (massimo.gion@aulss3.veneto.it)
Anne Rutjes (anne.rutjes@isp.unibe.ch)

Version: 1 Date: 17 May 2019

Author’s response to reviews:

Per Olav Vandvik (Reviewer 1)

Comment 1. I read this manuscript with great interest as I have been concerned about the trustworthiness of oncology guidelines, in particular given their lacking inclusion of systematic reviews to inform their recommendations, and availability of fully informative evidence summaries to inform patients and health care professionals about treatment decisions.

The paper contains some interesting findings and is generally well performed in terms of methods for their systematic review. A catchy finding is that only 28% could document appropriate systematic searches together with other findings to further document limitations and some regression analyses pointing to specific problems and trends.

Reply to comment 1:

We thank the reviewer for his positive general comment.

Question 1. The main findings suggest that the majority of oncology guidelines can not be classified as trustworthy according to standards, unless these by other mechanisms have found a relevant and high-quality up-to-date systematic review to inform recommendations, which precludes the need to do a systematic literature search from start.

Whereas this is unlikely it points to the major limitation of the paper which would warrant a major revision in my view, to make it more relevant and newsworthy for the readership of BMJ
MRM. The major limitation in my view relates to the very narrow objective of the paper, focusing solely of systematic searches for literature rather than the appropriate use of systematic reviews to inform recommendations. Although the paper, as mentioned, nicely demonstrates that only 28% performed adequate searches according to the defined standards it does not describe to what extent the guidelines appropriately synthesised the identified studies and ended up using systematic reviews to inform recommendations. This is in my view the key question that this paper ideally should address, which would warrant a major revision. It would require that authors re-assess the guidelines and document for each recommendation to what extent it was based on a systematic review of sufficient quality and to what extent this was communicated to the readers, in terms of evidence summaries providing estimates of effect on benefits and harms and reported quality of evidence. This could be mapped against IOM standards or AGREE II for the evidence part of guidelines. Another limitation is the search which ended in 2015, questioning the extent to which the findings are applicable in 2019, assuming that also oncology guideline organisations may have improved over time.

Reply to question 1:

Clinical practice guidelines typically address multiple questions, so that by default, several systematic reviews are required to address those questions. The situation sketched by the reviewer is theoretically correct; however, in practice, a systematic search to identify systematic reviews or other scientific evidence is always indicated. This is the reason why, although the suggestions of the reviewer are interesting, they go beyond the scope of our review. We first want to describe the essence, to describe fraction of guidance documents adopting systematic review questions. To address the objective of the reviewer, a different design would be required, it would require at least in part qualitative research using, for example interviews with guideline panel members (chairs) and observations of the guideline process. In addition, the suggested focus would require restriction to inclusion of higher quality guidance documents with high quality of reporting. Overall, the objective and subsequently the design of our study differs from the reviewers suggestion importantly, and we prefer to keep the design as used, which fitted the narrow objective well. As shown by our analyses, no substantial improvement was detected over time, so that our findings are still deemed to be relevant to current standards.

Comment 2. I would be happy to provide more detailed comments to the paper but will wait to see whether the journal agrees that a more comprehensive scope is needed which then would warrant a major revision. If the journal would like to accept the paper with minor revisions I could add some specific details in the next round.

Reply to comment 2:

We look forward receiving more detailed suggestions for improvement.
Joshua Pink (Reviewer 2)

Question 2. The title of the paper should probably reflect the fact that you only included oncology guidelines in your review - you note in your discussion that "it remains to be investigated if our results can be transferred to other healthcare areas", so the title may be too general.

Reply to question 2:

We adapted our title in: “Insufficient uptake of systematic search methods in oncological clinical practice guideline: a systematic review.”

Question 3. On P4, the paragraph starting on line 90 needs a citation as to where this argument comes from. I don't object to this classification, but I think there definitely are people who would regard their documents as CPDs when you are saying they aren't.

Reply to question 3:

We now reference to https://www.ahrq.gov/gam/summaries/inclusion-criteria/index.html.

Question 4. On P5, you should probably give a reason for the first three bullet points of the excluded list, in particular whether this was done simply to manage the volume of work, or whether you had a hypothesis those types of guidelines would be fundamentally different in some way in their likelihood of using systematic searching methods.

Reply to question 4:

We thank the referee for the comment. The reasons for the exclusion are now more clearly explained. In fact, the study is collateral to a national project aimed at synthesizing recommendations on traditional circulating tumor markers (this point is stated in the introduction, page 4, line 98). Therefore, we excluded guidelines concerning (i) malignancies in which traditional circulating tumor markers are not considered (brain tumors, soft tissue tumors, musculoskeletal tumors, hematological malignancies, non-melanoma skin cancers, hereditary/familial cancers) and (ii) malignancies in which traditional circulating tumor markers haven’t any role in risk assessment or in the management of specific patients (documents focusing only on screening, prevention or palliative care, documents focused on a specific subgroup of patients). Accordingly, the manuscript has been amended (and simplified) as follows:
Introduction (page 4 line 98)

“The objective of this study, that is collateral to a national project aiming at synthesizing recommendations on traditional circulating tumor markers in solid tumors was to assess the use of systematic search methods in CPGs in oncological field as a test context framework, and to determine which context variables were associated with adequate use.”

Methods (from page 5 line 127 to page 6 line 133)

“The following documents were excluded from the analysis:

• guidelines concerning malignancies in which traditional circulating tumor markers are not considered (e.g. brain tumors, soft tissue tumors, musculoskeletal tumors, non-melanoma skin cancers, hematological malignancies, hereditary/familial cancers);

• guidelines concerning malignancies in which traditional circulating tumor markers haven’t any role in risk assessment or in the management (documents focusing only on screening, prevention, palliative care; documents focused on specific subgroups of patients, such as pregnant women, children, adolescents, homeless people).”

Question 5. On P6 line 137, I presume that even though you included all instances of the report, it was only one data point in your logistic regression, and that is probably worth stating explicitly.

Reply to question 5:

We agree and adapted the text in (page 6 lines 144-148) “Some reports were identified as multiple reports belonging to an unique guidance document if they were either solicited by the same organization/producer or produced during the same consensus meeting or if they were explicitly cross-referenced in the main guidance document. In these cases all identified multiple reports were included to allow for a full appreciation of the guideline but they contributed together as one data-point in our logistic analyses.”

Question 6. On P7 line 163, it would useful to explain why you chose this definition. It isn't unreasonable, but obviously there are others you could use (some people would argue you need to use multiple databases to be systematic), so a justification for this particular one would be good.

Reply to question 6:
We derived it from Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138(8): 697-703. Quote from that paper “Ideally, such strategies include the use of one or more bibliographic databases (including a specification of the key words and other aspects of the search strategies39), a search for reports that cite the important papers found through a database such as the Science Citation Index, perusal of the references of all the relevant papers found and personal communication with investigators or organizations active in the area being reviewed (to make sure important published papers have not been missed and particularly to look for methodologically adequate studies that have not been published).”


A similar definition was used by NICE in 2004, please see https://www.ncbi.nlm.nih.gov/books/NBK49325/.

We now embedded these citations in the manuscript. In lack of broad consensus in the evidence synthesis field, we did not opt for a stricter definition, which would have resulted in even less articles meeting our requirements.

Question 7. On P8, your description of model 2 does not seem to match the results in table 3. I was expecting from this methods description for you to present one model (SR vs SRinsDT and SRnoDT) but in fact you seem to report two models for SR vs SRinsDT and SR vs SRnoDT separately. It would be good to either correct that or, if I am misinterpreting what you have done, make what you have done clearer.

Reply to question 7:

We thank the reviewer for the comment. We add the following sentences in the text, page 9 lines 209-210: “We presented the results of Model 2 into two separate analyses, (SR vs SRinsDT) and (SR vs SRnoDT), respectively.”

Question 8. It would seem useful to include an additional analysis in your paper, namely studies using appropriate systematic methods vs those not doing so (for whatever reasons) - so SR vs SRinsDT+SRnoDT+noSR+noLA. The conclusions you are trying to draw seem ultimately about what correlated with people using appropriate methods vs not, but this is an analysis you don't present (you only have people using some systematic versus not).

Reply to question 8:
The rationale of the analyses reported in the manuscript was to compare two different sets of approaches adopted by guideline developer panels: (i) documents prepared with the intention to use a systematic revision (subsets: SR, SRTinsDT, SRnoDT); (ii) documents prepared without the intention to use a systematic revision (subsets: noSR, noLA). Within the set of documents prepared with the intention to use a SR, we performed an additional comparison between SR appropriately performed (SR) and SR still requiring improvement (SRTinsDT and SRnoDT).

We performed the additional analysis suggested by the reviewer (results reported in Additional file FOR REVIEWERS, only for reviewer information). Although similar effects were found in this post-hoc, the evidence for effects of scope and continent were weaker.

As these analyses were not protocol-based and do not respond to the above mentioned rationale of comparison between homogeneous sets, we prefer not address them in the manuscript.

Question 9. I presume in your list of included guidelines you will have had multiple guideline from the same organization, and in that case it would seem appropriate to include that as a clustering variable in your regression analyses. Perhaps if you have no more than 2 or 3 guidelines from any given organization it might not make much difference, but if there are some organizations with a significant number of included guidelines it could.

Reply to question 9:

We agree with the comment of the reviewer and we rerun the analyses, taking into account clustering. We thus adapted confidence intervals in Table 2 and 3 and in the whole text. We add the following sentences in the text, page 9 lines 211-212: “Documents produced from and published by the same organization, were include as a clustering variable in the regression analyses to obtain robust variances.”

Question 10. I think you discussion does need to have a bit more on what the "North American" finding means. In particular, it could be geographical, but presumably it would also be guidelines from high-income countries using more systematic methods (since the guideline producers presumably have more money), with North America a proxy for that since both the USA and Canada are high income, whereas all your other continents could contain a mix of higher and lower income countries (I don't know if they do in practice).

Reply to question 10:

The referee comment is appropriate, as North America is a proxy for USA and Canada, which are high-income countries. However, guidelines produced by low income countries was really rare in our survey. The positive associations found between declared use of systematic search
methods and guideline production in North America seems not due to the fact that both the USA and Canada are high income countries. In fact, 551 out of 590 examined guidance documents are produced by high income countries and guidance documents from two other continents (Europe and Oceania) are all produced in high income countries. Moreover, only 18 out of 590 (3%) guidance documents were produced in lower income economies countries.

This point has been reported in the discussion (page 13 lines 310-314).

Question 11. I think there does need to be an additional supplementary appendix giving a list of all the included guidelines. Ideally I would like to see how each individual guideline has been classified as well, but just a list would still be better than nothing.

Reply to question 11:

Thank for the suggestion. One more table was added (Additional File n. 2).