APM ANNALS OF PALLIATIVE MEDICINE
AN OPEN ACCESS JOURNAL FOR HIGH-QUALITY RESEARCH IN PALLIATIVE MEDICINE
JOURNAL CITATION REPORTS IMPACT FACTOR 1.681 © 2020 Clarivate Analytics

Peer Review File

Article information: https://dx.doi.org/10.21037/apm-21-712

Reviewer Comments

Generally, the manuscript needs a major revision. See my detailed comments below:

In general:

Comment 1: Why is the text full of direct quotations (but then sometimes several references are cited), e.g. lines 76-78? These should only be used very sparingly and to emphasize a particular statement! Please revise the entire text accordingly!

Reply 1: Thank you for this valuable feedback. We have revised the full text according to your suggestion.

Changes in the text: We have reduced the use of direct quotations in the full text.

Comment 2: I am not a native English speaker, but it seems compelling to me that the entire text needs a linguistic English revision. Also use tools like "grammarly".

Reply 2: Thank you for this valuable feedback. We have revised the full text according to your suggestion.

Changes in the text: We have further revised the manuscript for style and language.

Abstract:

Comment 3: Line 48: I completely agree here, but what the authors conclude has already been implemented. For example, the checklist in the included Langer et al. 2018 study is based on guidelines that were directly adopted into the checklist, and thus have "high-quality evidence", because the guidelines are generally based on all the measures that the authors suggest here (steps 1-5 in the results).

Reply 3: Thank you. We have modified according to your suggestion as follows:

Changes in the text: SPs have been proven to be an effective method to assess performance in practice. There are still some deficiencies in the developing of case-specific checklists using SPs. To ensure the validity and reliability of checklists, the development processes need to be more standardized and procedural. (see Page 3, line 44-47).

Introduction:

Comment 4: Line 66-68: Fully agree!

Reply 4: Thank you.

Changes in the text: None.

Comment 5: Line 75: "for us": In a scientific text, this should be avoided. Can be omitted without replacement.

Reply 5: Thank you. We have deleted such a description in the full text according to your suggestion, as follows:

Changes in the text: For these reasons, it is necessary to use other means of obtaining data for the evaluation of quality. (see Page 4-5, line 68-70).

Comment 6: Line 80-82: Not all sources of this review refer to physicians, e.g. [20] Zolezzi et al. 2019 to pharmacies. Please revise accordingly!

Reply 6: Thank you for this valuable feedback. There may be some ambiguity in our original description. Therefore, we have modified the relevant description in the full text according to your suggestion.

Changes in the text: We have replaced the term “physicians” in the full text with “medical workers”.

Comment 7: Line 87: What are these gaps? These should be described in a little more detail.

Reply 7: Thank you. We have added more description of these gaps as follows:

Changes in the text: However, gaps in the reporting of checklist development methods remain to a certain extent, including the failure to base on high-quality evidence, and lack of validation by multidisciplinary experts, etc. (see Page 5, line 82-84)

Methods:

Comment 8: Line 102: Very few keywords and no truncation were used. This should be mentioned as a limitation.

Reply 8: Thank you for this valuable feedback. When searching English databases, we have used truncation, such as “standardized patient*” and “simulated patient*”. We have added a statement on the index bias in the limitation section according to your suggestion. We have modified the manuscript as follows:

Changes in the text: This study also has some limitations. Due to the few SP-related keywords in indexing, some articles may not be identified. Another limitation of our

study was that it only included articles published in English and Chinese, which could introduce publication bias (see Page 16, line 291-293).

Comment 9: Line 107: The behaviour of physicians is to be assessed, so why were pharmacy-related studies (e.g. Zolezzi et al. 2019 and Langer et al. 2018) included?
Reply 9: Thank you for your question. The purpose of this study is to develop an evidenced-based quality assessment checklist in primary health care settings. In the process of searching and screening relevant literatures, we included all the studies using standardized patients to assess the medical workers' performance, and are not limited to physicians. Moreover, if only studies on physicians are included, the number of studies meet the criteria would be small. There may be some ambiguity in our original description. Therefore, we have modified the relevant description in the full text.
Changes in the text: We have replaced the term "physicians" in the full text with "medical workers".

Comment 10: Line 123-124: Please describe in more detail the "6S"-Model.
Reply 10: Thank you. We have added more description of the "6S" model as follows:
Changes in the text: By summarizing the methodological information of the checklist development, that is, we would take some procedures as key steps if more studies have adopted them. Meanwhile, according to the recommendations of the WHO Handbook for Guideline Development (38) and the "6S" model, which is often used to identity the best evidence for a clinical issue (39), we further verified and supplemented the initial key steps, and proposed the final key steps eventually. The "6S" pyramid model was proposed by Dicenso A et al. in 2009, which includes six levels from top to bottom: systems (e.g. computerized decision support systems), summaries (e.g. evidenced-based clinical practice guidelines), synopses of syntheses (e.g. ACP Journal Club), syntheses (e.g. systematic reviews), synopses of studies (e.g. evidenced-based abstraction journals) and studies (e.g. original articles published in journals). Clinical evidence can be indexed sequentially until reliable and valid evidence is obtained at a certain level (39) (see Page 7-8, line 119-130).

Results:
Comment 11: Lines 150-156: Are the authors sure that this information is actually always correct in individual cases?
For example Zolezzi et al: Was it actually reported that pharmacy graduates were the

developers of the checklist? Analogous to Langer et al.: Was it actually reported that an expert panel was used?

I think not.

The authors should therefore review all included studies again with regard to the individual criteria.

Reply 11: Thank you for this valuable feedback. We have reviewed and corrected all the information of included studies again according to your suggestion, and there are indeed some misunderstandings before. At the same time, the information in the full text was modified accordingly.

Changes in the text: We have corrected and updated the information in the results section and Table 1.

Comment 12: Lines 157-163: In my opinion, the authors overlook the possibility that the guidelines are already available in a form that can be used to create a checklist without further steps (e.g. Langer et al. 2018, hence the unreported expert consensus). If they are official guidelines, they are already based on literature and an expert consensus.

Therefore, it also seems questionable whether the checklist development formulated by the authors following this paragraph always has to be done in this way.

Reply 12: Thank you for this valuable feedback. We agree with you very much. According to the WHO Handbook for Guideline Development, one of the essential steps is that the expert group achieve consensus on recommendations based on the evidence proposed by methodologists, while taking into account preferences and values of patients, as well as economic cost of interventions, etc. So, if there are official guidelines, it's probably based on literature and expert consensus. However, on the one hand, the quality of these guidelines is uncertain, so we suggest that only high-quality guidelines should be included after comprehensive searching and assessment. On the other hand, since the method proposed in this study is mainly applicable to primary health care settings, considering that some recommendations in the published guidelines may not be applicable to primary health care settings, we propose to further ensure the applicability and rationality of the final checklist through expert consensus.

Changes in the text: None.

Comment 13: Line 165: In general, the question arises whether and in what way the checklist development on the following pages is based on the results of the review?

How were the findings of the review (table 1) transferred to the checklist development (table 2)?

Reply 13: Thank you for your question. The findings of the review (table 1) focused on three main aspects: who developed the checklist, what development procedure was used, and whether the development process was evidence-based or generated through consensus procedures. By summarizing these findings, that is, if more studies adopt a certain step, we believe that it's likely to be a key step. Meanwhile, we also refer to the recommendations of the WHO Handbook for Guideline Development and the "6S" model to further verify and supplement the key steps, and proposed the procedure for checklist development (table 2) eventually. We have supplemented the method in the manuscript as follows:

Changes in the text: By summarizing the methodological information of the checklist development, that is, we would take some procedures as key steps if more studies have adopted them. Meanwhile, according to the recommendations of the WHO Handbook for Guideline Development (38) and the "6S" model, which is often used to identity the best evidence for a clinical issue (39), we further verified and supplemented the initial key steps, and proposed the final key steps eventually. (see Page 7-8, line 119-124).

Comment 14: Lines 204-207: See also my comment on lines 157-163: However, this is exactly the decisive point: if official guidelines of a relevant professional association are available (what do the authors mean by "high-quality guidelines"?), then the further steps are not necessary, especially if these guidelines are designed in such a way that the checklist can be obtained directly from them or if, for example, a checklist is available in the appendix to the guidelines (e.g. in Germany for the dispensing of oral emergency contraceptives in community pharmacies). Please revise accordingly, the 1 sentence in line 204-205 is not enough.

Reply 14: Thank you for this valuable feedback. Although there may be official guidelines, the quality of guidelines can be extremely variable and some often fall short of basic standards, so we suggest to evaluate the quality. The AGREE instrument recommended in this study is a tool that assesses the methodological rigour and transparency in which a guideline is developed. It consists of 23 key items organized within 6 domains followed by 2 global rating items ("Overall Assessment"). Each domain captures a unique dimension of guideline quality, and the quality score is calculated for each of the six AGREE II domains. After using this tool

for evaluation, we can distinguish those high-quality guidelines. We have supplemented the original text according to your suggestion as follows:

Changes in the text: Inclusion criteria for the literature should contain eligible guidelines and consensus, systematic reviews, and studies on diagnostic and treatment for target diseases sequentially. Followed by the quality evaluation of included studies, as shown in the "quality evaluation and inclusion of final literature" below. Literature can be retrieved according to the general principles: If eligible high-quality guidelines can be retrieved, then there is no need to search consensus, systematic reviews or the lower-level literature unless the upper-level literature cannot answer the diagnostic and treatment problems of target diseases. Otherwise, it is suggested to continue to search for other levels of eligible studies in order until the diagnosis and treatment of the target disease can be answered (see Page 11-12, line 207-215).

Comment 15: Line 215: Why is it necessary to evaluate guidelines (They are usually created by a process that the authors describe on pages 9ff.)?

Reply 15: Thank you. Since there may be many guidelines related to the target disease, but the quality of these guidelines is generally various, so we need to evaluate guidelines to exclude some low-quality ones. Only on the basis of evidence of high-quality, can we get a more scientific and accurate checklist.

Changes in the text: None.

Comment 16: Line 216: Please explain more precisely: Is it meant here that a quality evaluation should lead to exclusion and then the "inclusion criteria" apply to the remaining studies?

Reply 16: Thank you for this valuable feedback. What we mean is that after screening out the literatures that initially meet the inclusion criteria, they should be evaluated for the methodological quality, and then exclude those with low-quality from the included studies. In other words, only the high-quality studies that meet the inclusion criteria will be included eventually. We have revised the description of this part according to your suggestion, as follows:

Changes in the text: Appraisal of Guidelines for REsearch & Evaluation II (AGREE II) (53), A MeaSorement Tool to Assess systematic Reviews (AMSTAR) (54) and QUality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (55) should be used for evaluating the methodological quality of the preliminary included guidelines, systematic reviews and diagnostic tests, respectively. Although some studies may meet the initial "inclusion criteria", if they are found to be of low-quality or there are

large defects, they should be eventually excluded and continue to search for other high- quality studies that meet the requirements (see Page 12-13, line 222-228).

Comment 17: Lines 226-227: This sentence would be better placed with the description of group B in lines 177-179.

Reply 17: Thank you. We have modified according to your suggestion as follows:

Changes in the text: Group B, the clinical expert consensus group (six to seven experts) should include specialists in all relevant areas of the target disease, whose education, experience, region and gender also need to be considered. They are responsible for discussing and reaching consensus on the checklist items proposed by Group A (see Page 10, line 179-182).

Discussion:

Comment 18: Line 242: See also my comment on line 107: It is about the performance of physicians, but why were studies included that deal with other service providers (e.g. community pharmacies)?

Reply 18: Thank you for your question. The purpose of this study is to develop an evidenced-based quality assessment checklist in primary health care settings. In the process of searching and screening relevant literatures, we included all the studies using standardized patient to assess the medical workers' performance, and are not limited to physicians. Moreover, if only studies on physicians are included, the number of studies meet the criteria would be small. There may be some ambiguity in our original description. Therefore, we have modified the relevant description in the full text.

Changes in the text: We have replaced the term "physicians" in the full text with "medical workers".

Comment 19: Lines 265-269: See my previous comments on this topic: What if guidelines from a professional society are already available (usually developed exactly as the authors suggest)?

Reply 19: Thank you for your question. Even if there are relevant professional guidelines, the method we propose is still applicable. On the one hand, the quality of these guidelines is uncertain, so we suggest to evaluate the quality by AGREE II instrument. After evaluation, we can distinguish those high-quality guidelines. It is noteworthy that in the step of "Selecting and evaluating relevant references", because there are already relevant high-quality guidelines, there is no need to search for other

JOURNAL CITATION REPORTS IMPACT FACTOR 1.681 © 2020 Clarivate Analytics

evidence like systematic reviews, diagnostic tests etc. On the other hand, since the method proposed in this study is mainly applicable to primary health care, considering that some recommendations in the published guidelines may not be applicable to primary health care settings, we propose to further ensure the applicability and rationality of the final checklist through expert consensus and pre-test.

Changes in the text: None.

Comment 20: Lines 279-280: Yes, but not only how to weight, but also whether to weight at all.

Reply 20: Thank you. We have modified according to your suggestion as follows:

Changes in the text: Future research needs to explore whether and how to weight the items of the checklist (see Page 15-16, line 288-289).