Author’s response to reviews

Title: Guideline-based quality indicators – a systematic comparison of German and international clinical practice guidelines

Authors:

Monika Becker (monika.becker@uni-wh.de)
Jessica Breuing (jessica.breuing@uni-wh.de)
Monika Nothacker (nothacker@awmf.org)
Stefanie Deckert (stefanie.deckert@uniklinikum-dresden.de)
Marie Brombach (marie.brombach@uniklinikum-dresden.de)
Jochen Schmitt (jochen.schmitt@uniklinikum-dresden.de)
Edmund Neugebauer (edmund.neugebauer@mhb-fontane.de)
Dawid Pieper (dawid.pieper@uni-wh.de)

Version: 1 Date: 05 Jun 2019

Author’s response to reviews:

We thank the reviewers for their comments and advice on our manuscript! We studied the comments carefully and revised the manuscript accordingly.

Reviewer #1:

This is a well-written and well-structured paper describing and comparing quality indicators (QIs) from Germany and other countries (Belgium, Canada, US, UK) derived from evidence-based clinical practice guidelines (CPGs). This study is part of a German project on methods of translating guideline recommendations into QIs, which is also useful for the international audience.
The objective of this study was to compare guideline-based QIs identified in German CPGs, as well as their underlying methodological approaches, with those of international CPGs on related topics. Many relevant data on CPGs and QIs were collected, presented, and compared using descriptive statistics. Although hypotheses were suggested, these were not statistically tested.

Response of the authors: Indeed, in our study, we had the hypothesis that in many cases QIs from German S3-CPGs do not correspond with QIs of international CPGs. However, we decided not to perform statistical testing due to the heterogeneous nature of the CPGs. They varied greatly for example time period of literature searches and publication date, developing organization and health care context as well as in the scope. Thus, we think that statistical testing would not increase the interpretability of our results. We added the following sentences under the discussion section: “This result confirms our hypothesis that in many cases, QIs from German S3-CPGs do not correspond with QIs of international CPGs on related topics.” and “Although we suggested a hypothesis, we decided not to perform statistical testing due to the heterogeneous nature of the CPGs. They varied greatly for example in time period of literature searches, publication dates, developing organization, health care context as well as in the scope.”

One major comment is that the authors collected CPGs on same topics and then directly compared the QIs of CPG pairs. The step of identifying the individual recommendations within the CPGs and assessing their comparability was skipped. Not surprisingly, most of the QIs (83%) were not comparable as the individual underlying recommendations were not comparable. For comparing QIs, the individual recommendations should be distracted from the CPGs first and then the QIs can be compared. The authors should acknowledge this in the Discussion section. This will not affect the conclusion that information on the methodological approach of developing QIs is often lacking.

Response of the authors: Indeed, our approach was first to collect QIs and then to examine how the QIs were linked with recommendations. Although it might be expected that QIs reported in CPGs should be directly linked with recommendations, we found that about half of the QIs reported in international CPGs were not based explicitly on guideline recommendations. We only considered QIs on clinical topics that were addressed in both CPGs of a CPG pair.
However, we agree that this point should be acknowledged in the discussion. In the section “reasons for differences in QIs”, we added the following paragraph: “Furthermore, although we compared only QIs on clinical topics that were addressed in both CPGs of a CPG pair, several recommendations of the German S3-CPGs and the related international CPGs varied to some extent in content and definitions. Most of the recommendations reported in international and German S3-CPGs were not inconsistent, but had a different focus or depth of detail. For example, the German S3-CPG “Type 2 diabetes training” recommended to offer a structured education program, whereas the international CPG conducted by ICSI on “Diagnosis and Management of Type 2 Diabetes Mellitus in Adults” comprised a specific recommendation of nutrition therapy. Nutrition therapy was also considered in the particular German S3-CPG within the explanatory text. However, no specific recommendation on nutrition was made. Further, there were other cases where both CPGs of a CPG pair comprised comparable recommendations, but only in one of the CPGs a QI was derived from the recommendation(s).” Further, we added under limitations “Finally, the interpretability of our results might be limited as we compared the QIs on clinical topics that were addressed in both CPGs of a CPG pair directly, rather than at the recommendation level. As noted above, although the CPG pairs addressed the same clinical topics, the recommendations varied to some extent and, in some cases, resulted in QIs that were not comparable. However, it should be noted that only about half of the QIs reported in international CPGs were based explicitly on guideline recommendations. The underlying approaches for generating such QIs were not reported in sufficient detail.”

One minor comment is that reflection on the quality appraisal scores of the CPGs is missing. The domain scores vary from 48 to 83%. Did the authors consider to use a threshold, for instance excluding CPGs with a score lower than 50 or 60% arguing that those are less evidence-based? And is a high scoring CPG related to better description of the methods of developing the QIs?

Response of the authors: We did not consider to use a threshold for excluding CPGs. Although it is assumed that the degree of credibility of QIs is associated with the methodological quality of CPGs, there is lacking evidence for that. This might be evaluated in future research projects.

However, we agree that this point should be mentioned. We added the following in the discussion section: “The quality appraisal score for the domain “Methodological Rigor of Development” ranged from 50% to 83% and from 48% to 83% in the German S3-CPGs and international CPGs, respectively. High scores were not inevitably related to better description of the methods of developing the QIs or better reporting of QIs. Although it is assumed that the degree of credibility of QIs is associated with the methodological quality of CPGs, the evidence for this is lacking so far.”
Reviewer #2:

Thank you for giving me the opportunity to read and comment on the manuscript entitled "Guideline-based quality indicators - a systematic comparison of German and international clinical practice guidelines"

This manuscript aimed to compare guideline-based quality indicators of German and international clinical practice guidelines and their underlying methodological approaches.

Overall, it tackles an important topic. However, a very important methodological limitation is that the manuscript was submitted in the category of systematic reviews. Yet, selection and extraction of studies are done by one reviewer rather than two independent reviewers. This is an important detail that could be mentioned in the Abstract and the limitation section. So, for the type of design, authors might think about the possibility to indicate this study as a 'rapid review'.

Response of the authors: Title- and full-text screening as well as the extraction by two reviewers would have been certainly the ideal case. However, we decided to perform the selection of guidelines only by one reviewer due to the large number of hits by diverse searches as well as the nature of guideline searches. We consider this pragmatic approach in the area of guideline searches to be acceptable, particularly because of the very low level of complexity regarding inclusion criteria in our study. Also, the extraction was conducted by one reviewer and checked by another. However, this should be appropriate as it corresponds to a recent methodological guide regarding systematic reviews of CPGs (Johnston, A., et al. (2019) "Systematic reviews of clinical practice guidelines: a methodological guide." J Clin Epidemiol 108: 64-76.).

However, we agree with you that this should be mentioned in the abstract and the limitation section. Accordingly, we added the following information in the limitation section:

“Furthermore, potential limitations arise from the fact that both the selection of CPGs and data extraction were performed by only one reviewer and checked by another. This pragmatic approach was chosen because of the large number of hits obtained by the diverse searches, as well as the low level of complexity regarding inclusion criteria in our study. Moreover, the data extraction is in agreement with a recent methodological guide on systematic reviews of CPGs [Johnston et al. 2019].”

In the abstract, we added the sentence “The selection and extraction of CPGs were conducted by one reviewer and checked by another.”
We would rather abstain from labeling our study as “a rapid review” as we didn’t designate our study explicitly as systematic review and indicated under methods that our study “…did not fulfil all requirements related to a systematic review.”

Finally, in the Abstract section, I would cut down the introduction and conclusion length and provide more information under methods (e.g., study design, inclusion criteria, selection process) and results. The abstract could be reviewed; it could be inappropriate to start a sentence with a number.

Response of the authors: We have revised the abstract. We have shortened the introduction and conclusion and added information under methods and results. We rewrote sentences that started with a number in the abstract and rewrote some sentences in the main text under 3.2 Characteristics of guideline-based QIs und 3.4 Methods for the development of QIs, accordingly.