Author's response to reviews

Title: Evidence-based guidelines in the assessment of work disability: an international survey and a comparison of quality of development according to AGREE in Germany and the Netherlands.

Authors:

Wout E.L. de Boer (wout.deboer@tno.nl)
David J Bruinvels (d.bruinveld@vumc.nl)
Arie M Rijkenberg (A_M_Rijkenberg@lycos.at)
Peter Donceel (peter.donceel@med.kuleuven.be)
Han J.R. Anema (h.anema@vumc.nl)

Version: 3 Date: 21 August 2009

Author's response to reviews: see over
**Author’s response to reviewer’s comments**
The authors thank the reviewers for their extensive comments. It enables us to clarify many points that were addressed in a too implicit manner. At several places the reviewers indicate mistakes that should not have occurred, like not explaining abbreviations, for which we apologize. The authors are pleased to note that the reviewers consider the article of importance to those concerned. We summarized the comments and our handling of it, below.

**Response to Dr Burgers:**

*Major comments*

+ It is unclear why these Dutch and German guidelines were chosen. Because of the healthcare systems?
R: This choice was not because of the healthcare systems. We clarified now that the selection of disease oriented guidelines reduced the number of potential countries, and that comparable pathology reduced the number of potential guidelines.
+ There is no discussion on why the number of guidelines is so low in other countries involved in EUMASS. The relevance of the paper would increase if more information is provided about the social and political context of SIM in general.
R: we hypothesize in the Impact section that it is a matter of time before guidelines will turn up in other countries too. But we did not study that. The context of SIM and how it is different from clinical medicine is now elaborated in the Introduction section.
+ The second research question ‘to what extent are these guidelines evidence-based’ should be changed into ‘what is the quality of the guidelines in SIM’?
R: We thank dr Burgers for this suggestion and we agree that that is more precise with regard to AGREE; we changed this.
+ Language should be checked and improved by a native English speaker.
R: This had been done with the first draft and we had it done for this second draft as well.

*Minor comments:*

+ The abbreviation EUMASS should be explained in the paper.
R: This is done.
+ Conclusion in the abstract is too long compared to other sections in the abstract.
R: This was reduced, focusing on key messages.
+ The structure in the article could be clarified by adding headings.
R: We did so and thereby stayed close to the research questions.
+ Clarification is needed why these four countries and not other countries were visited.
R: on the basis of reporting to use guidelines. Other countries did not report this. This is clarified in the article.
+ The term ‘legal guidelines’ could be replaced by ‘process oriented guidelines’.
R: we did change that
+ Further characteristics of the guidelines could be added to table 1.
R: we added number of pages and references, as described in Burger et al 2002.
+ The term disease oriented is not used consistently in the text, with medical, disease-oriented, and diagnosis-oriented.
R: we opted for disease oriented and made that consistent throughout the text
+ Standard deviation or range could also be provided.
R: This suggestion we did not follow as we believe it not to yield more relevant information for this article: there were only two appraisers and the Kappa’s indicate the initial agreement.
+ No quantitative data are provided on Kappa’s.
R: As dr Burgers indicates these are not useful for this paper. So we did not change that.
+ Two experts (one for the Netherlands and one from Germany) were asked to comment on the AGREE scores in stead of scoring themselves.
R: We thought it easier for them rather than appraise the guidelines themselves. Doing as dr Burgers suggested might have increased the reliability of the scores but on the other hand they might have been biased. It is not unusual to ask for comments after scoring, e.g. in systematic reviews. Anyway that was not the way the project was designed. So we can not change that.
+ Initial agreement on scoring German guidelines was smaller than the Dutch ones. An explanation is needed.
R: this was mainly due to translation differences and different ideas of the German context. After discussion of the appraisers they agreed.
+ The limitations of the AGREE are not mentioned, such as that it particularly measures the methodology and reporting and not the clinical validity of the recommendations (Burgers, Clin Chem 2006).
R: we added this aspect in the introduction section. It was mentioned in the impact section already.
+ Another limitation is that only two reviewers appraised the guidelines. The AGREE Instrument recommends four appraisers.
R: We agree that four appraisers are usual measuring the AGREE criteria. The purpose of this paper was however to explore the quality of disease oriented guidelines in Insurance Medicine as is done in e.g. systematic reviews. For this reason, two independent appraisers were foreseen and a third independent reviewer in case of disagreement. Two appraisers are usual for a quality assessment in a systematic review. The two first reviewers reached full consensus and so the third appraiser did not come into action. Involved experts commented on this scoring. The reliability of this procedure seems acceptable for the purpose of this paper.
+ More discussion is needed on the impact of this study on other countries than only Germany and the Netherlands.
R: This aspect is addressed now in the discussion section.

Response to Dr Kroll

Major comments:

1
+ The authors could add a bit more background details regarding the scope of the problem and be more explicit in terms of the geographic extent of their work.
R: We specified the background and clarified the geographical extent in Introduction of the article.
+ The line of argumentation could be strengthened by focusing more closely at the social insurance context that constitutes the ground for their comparisons.
R: we clarified that the medical assessments are relatively comparable between social insurance systems.
+ Statements like ‘in most compensation systems...’ on page 4 for example require a reference.
R: This is done at all places that required it.
+ In their discussion of evidence-based guidelines in occupational medicine they draw upon the situation in the Netherlands but it is unclear to what extent the introduction of these guidelines is also happening elsewhere.
R: we refer to occupational medicine in the US and the UK as well. Guidelines are being implemented there too. We are not aware of this in other countries. However we wanted to illustrate that guidelines and the AGREE instrument are not restricted to the clinical environment.
+ The statement ‘Having guidelines does not necessarily mean that quality is supported’ (p. 5) needs more elaboration. Quality of ‘what?’
R: this is specified now as ‘quality of medical work’.
+ At times it is difficult to discern, what the principal objective of the study is and in my view three issues are addressed in the paper but not dealt with in adequate or sufficient detail. (1) The examination to what extent guidelines are available for specific clinical conditions in determining work-related disability in various countries; (2) the lack of existing, reliable and valid quality appraisal tools in social insurance decisions about work-related disability; and (3) the aim to establish the psychometric and feasibility properties of a quality appraisal tool that has found application in other areas of clinical guideline evaluation (the AGREE tool).
R: The first issue is addressed by us and answered as far as we could in this project. The second issue is not addressed by us in this article. We take this from the literature as reason to study the first and third issue. The third issue is addressed although to a limited extent as few guidelines were really available.
+ There is no critical discussion of simply transferring and adapting this tool from one area of quality appraisal to another.
R: this is an important aspect that we pay attention to in the discussion section now. However, this is short as that goes beyond our article.
+ Acronyms such as AGREE, OHP, EUMASS, SIP etc. need to be spelled out when mentioned for the first time.
R: we checked and corrected this.

2
+ The sampling criteria should be made more explicit. The authors have used the EUMASS table to select countries, in which guidelines could be in use. What criteria were applied to the table to determine inclusion/non inclusion?
R: we clarified this and referred to an article about this that is submitted to Occupational Medicine.
+ They further state that they visited four countries. I assume that the fifth country was The Netherlands. However, this should be made more explicit.
R: this is rephrased.
+ Also, they say that ‘the status of guidelines was discussed’ (p. 6). What was specifically discussed?
R: this is specified.
+ I assume the authors mean quality of evidence when they are talking about the
‘evidence base of guidelines’ (p. 6). Please clarify.
R: we did.
+ On page 7, the authors describe situations where two researchers cannot reach agreement and the involvement of a third researcher. Could they give a bit more clarification about how the involvement of the third researcher led to resolving the impasse?
R: the third researcher was to decide on his own rating and the ratings of the two others. As the two first researchers reached agreement this did not happen.
+ Was the ‘expert’ chosen to comment on the findings involved in the development of the specific guidelines in question? They authors only said that these experts had been involved in the development of several guidelines. Please clarify.
R: we clarified the precise role of the experts in guideline development in the section where they appear for the first time.

3
+ There is a bit of confusion around the geographic scope of the results. In the methods section, the authors initially mention four countries, the Czech Republic, Germany, the United Kingdom and Switzerland. In discussion the inventory of disease-oriented medical guidelines to assess disability for work, they mention Germany, Ireland, Switzerland and The Netherlands (p. 8). Subsequently, the authors only compare guidelines from Germany and The Netherlands for four medical conditions.
R: this was made consistent and clarified in the methods and results sections.
+ On page 9, the authors report on the Kappa scores and state considerable differences between ratings for the Dutch and the German guidelines. What contributed to these differences? Did language differences play a role as the authors appear to be native Dutch speakers? Or what explanations could be given for the differences in appraisal? Greater detail is also necessary in terms of how agreement was finally achieved.
R: this has been specified further now in the results sections.
+ On page 11, the authors make a curious statement: “Stakeholder involvement was also reduced because the subject of patients’ involvement was controversial in the beginning as patients might be biased with regard to the recommendations.” Firstly, patients are not the only ‘stakeholders’. Other stakeholders include social insurance physicians, guideline developers, employers, etc. Secondly, it is doubtful that patients would be the only one who would demonstrate a bias in their judgment. Further down the author cite the view of the German expert who explained that the ‘development of guidelines had been new in Germany and started out of a need of the assessors within the Institution of Social Insurance, which explains the limited involvement of stakeholders.’ But does this not highlight an important point that the German expert seems to have missed: the key to meaningful development of guidelines is the participation of ALL stakeholders in the process. The failure to consult patients in the process can hardly be remedied at a later stage.
R: We agree with dr Kroll’s observation. The statement is a citation of the expert. It touches on the particular position the claimants hold in social insurance. This has been addressed in the Introduction and the Impact section now. In the Netherlands patients are involved now, in Germany the opinion seems to be unaltered. But these observations are beyond this article.
Further on page 11, it is said that: ‘The selection of evidence and the formulation of recommendations were carried out according to what the German ISI experts felt were the most important’. This appears to be a fairly ad hoc process. Surely, there must have been criteria to critically appraise the presented ‘evidence’. This should be made more explicit.

R: we checked this with the German expert. The participants in the development were all experts in their field and relied on their knowledge of the literature. There was no formal review process with explicit criteria.

Page 12: The authors suggest that ‘minor adaptations’ of the AGREE instrument were necessary. What did these adaptations look like? What are the likely implications for the integrity of the instrument?

R: we addressed this point with several text fragments on this in the discussion section now, referring for details to the additional file.

The authors do not provide any discussion of limitations except more mentioning that ‘the results are restricted to eight Dutch and German’ guidelines. Information could be added in terms of critically appraising the AGREE tool. What worked? What did not work?

R: we included a section on AGREE in the discussion section. Basically everything worked. The external validity is limited however as we were able to score guidelines of only two countries.

What are the implications of a comparison between two countries only for other countries? What about other medical conditions than the four selected for the comparison?

R: We added some more text on the scope of our results in the discussion section.

Minor essential comments:

The project is embedded in the literature. Since the principal comparison involves Dutch and German medical guidelines, it could be useful to elaborate a bit more on the similarities and differences of the social insurance situation with regard to employment of people with disabilities in those two countries.

R: we mentioned this aspect in the Impact section of the discussion but did not want to elaborate too much on it as we wanted to focus on the common aspect, the medical assessment of functional capacity.

The term ‘disability for work’ should be reworded as ‘work disability’.

R: we did follow this suggestion.

As in the main body of the text, the authors should be a bit more specific in the title and abstract terms of what constitutes their primary basis for comparison: the Dutch and German social insurance context.

R: we specified the title and the abstract to this end.