Reviewer's report

Title: Evidence-based guidelines in the assessment of disability for work: quality of development according to AGREE.

Version: 2 Date: 8 July 2009

Reviewer: Thilo Kroll

Reviewer's report:

1. Is the question posed by the authors well defined?

The authors set out to answer two principal descriptive research questions: (1) What guidelines have been developed in order to support the medical assessment of disability for work? (2) To what extent are these guidelines evidence-based?

The authors make a convincing case for the need of evidence-based assessments of guideline quality in work-related disability decisions. The authors could add a bit more background details regarding the scope of the problem (e.g. number of work-related disability cases, resulting cost implications) and be more explicit in terms of the geographic extent of their work. In the first sentence, they speak of the 'western world' but then they limit their work ultimately to comparing guidelines in the Netherlands with those in Germany. Social insurance systems vary tremendously across member states in the European Union and even more when considering the United States. The line of argumentation could be strengthened by focusing more closely at the social insurance context that constitutes the ground for their comparisons. Statements like 'in most compensation systems...' on page 4 for example require a reference.

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.

The statement ‘Having guidelines does not necessarily mean that quality is supported’ (p 5) needs more elaboration. Quality of ‘what’?

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 (or return to work? need for occupational adaptations?) 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). The acronym AGREE is not defined in the article, and needs to be spelled out and explained. And there is no critical discussion of simply transferring and adapting this tool from one area
of quality appraisal to another.

Acronyms such as AGREE, OHP, EUMASS, SIP etc. need to be spelled out when mentioned for the first time.

2. Are the methods appropriate and well described?

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? They further state that they visited four countries. I assume that the fifth country was The Netherlands. However, this should be made more explicit. Also, they say that ‘the status of guidelines was discussed’ (p. 6). What was specifically discussed? Was a formal appraisal undertaken? If so, how was this done?

I assume the authors mean quality of evidence when they are talking about the ‘evidence base of guidelines’ (p. 6). Please clarify.

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?

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.

3. Are the data sound?

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.

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.

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.

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.

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.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition? 5. Are the discussion and conclusions well balanced and adequately supported by the data?
6. Are limitations of the work clearly stated?

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

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?

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

The authors reference at least one publication that involved members of the team (deBoer et al., 2007). 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.

8. Do the title and abstract accurately convey what has been found?

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

9. Is the writing acceptable?

The writing is adequate. Areas in need of further clarification have been pointed out earlier.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Points made under 7. and 8. should be considered minor essential revisions.

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

Issues raised in points 1. through 6. should be considered compulsory as they constitute areas in which clarification is necessary.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declared that I have no competing interests.