

Reviewer's report

Title: Production and quality of clinical practice guidelines (CPG) in Argentina (1994-2004): a cross-sectional study.

Version: 1 **Date:** 20 February 2008

Reviewer: Jako Burgers

Reviewer's report:

Review Manuscript Production and quality of CPGs in Argentina

This is an interesting study with the aim to explore the production and quality of guidelines in Argentina. This is the second guideline appraisal study in a Latin American country, but on a larger scale than the first conducted in Brasil. Because of the overall low quality of the guidelines, there is a need for improvement of methods and national and probably international collaboration. The authors describe their findings in terms of representing a LMIC, which may increase the generalisability. Beyond Brasil, it is, however, unknown what happens in other countries. An international survey would be needed to explore this.

I have the following major comments:

- The guidelines were selected in two steps. In the second step 7 inclusion criteria were used. Yet, the authors did an extensive analysis of the 437 potential guidelines retrieved in the first step by categorising them according to the 'level of institution'. I think this is confusing as the study focusses on guidelines. They should limit the analysis to only those guidelines (n=144) that meet the inclusion criteria.
- Information about the names and number of organisations producing guidelines is lacking. I think this is necessary before categorising them. The authors could provide a list of organisations in a supplementary file.
- The AGREE scores are only discussed on the domain level. Information on item level is lacking. Particularly in case of low scores, it is necessary to know which items explain the low scores. The AGREE domains are aimed for summarising the items scores, but the items scores provide the most relevant information on quality. The authors should at least provide the item scores of the recent guidelines (not older than 5 years). The relation of the scores with other variables, such as the level of institution, can be analysed on the domain level. See Burgers et al. IJHTA, 2003.

Minor and discretionary comments

Introduction

- Suggested reference voor G-I-N is Ollenschlaeger et al, QSHC 2003.
- Ref 5 The AGREE Collaboration should be replaced by AGREE Research

Trust, www.agreetrust.org.

- A good guideline should ultimately (instead of eventually) lead to better patient outcomes.
- However, as this type of evidence is rarely available: unclear sentence. The authors confuse assessing quality with definition of quality. I would skip the sentence
- Shaneyfelts instrument is no ad hoc instrument but rigorously developed. However, it was not internationally validated.
- The AGREE collaboration published the results of the first international (!) project
- Last 3 sentences: Misuse of evidence. This should be replaced by information or studies

Methods

- I do not understand ref 15. The AGREE Instrument does not contain search strategies, so the translation should not include these either.
- Please provide information on how the appraisers were selected, more information on the training (one or half day workshop?) and how many guidelines (excluding the pCPGs) each appraisers assessed
- Please also provide the items of the AGREE Instrument in the supplementary file

Results

- See major comments: Guideline production should only include those guidelines selected after the second step.
- Figure 1: the term disease management is unclear. I suggest to split this in diagnosis and treatment/management.
- I would suggest to design one table on the characteristics of the guidelines, including the scope, year of publication and level of institution (with frequencies).
- Figure 3 is unclear. A table with mean domain scores, 95% BI and range of scores would be more informative. Please provide title, legend and explanation of abbreviations.
- See major comments: please also provide the item scores.

Discussion

- The item scores can be used for explanation the low quality as measured with the AGREE Instrument. Some low item scores can be explained by low resources.
- Are there differences between regional and national guideline production and how often have the concept of adaptation been used?
- How much would the lack of related documents have influenced the scores?
- The AGREE Instrument assesses the methodology and reporting but not the

cinical content. Suggested ref: Burgers, Clin Chem, 2006.

- An interesting document for discussion is the recently released IOM report [What really works in health care](#) about the need for a National Effectiveness program in the US, to tackle pluralism and low quality. So even in high income countries, low quality is an issue.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.