Reviewer's report

Title: Patient safety culture measurement in primary care. Clinimetric properties of SCOPE

Version: 1 Date: 15 April 2011

Reviewer: Barbara Hoffmann

Reviewer's report:

Thank you for this paper and the opportunity to comment on it. It reports on an important subject that had not been of much attention in the past: patient safety culture in general practice. It demonstrates the results of an important project that adds valuably to the scientific knowledge. It deserves publication with only minor, but essential revisions.

- Major Compulsory Revisions

None.

- Minor Essential Revisions

General comments:

1. Context: In the manuscript (and the title) the terms primary care and general practice are used interchangeably. However, the questionnaire deals with safety culture in general practices. Is primary care in the Netherlands just general practice? If the answer is no terms should be made clear.

2. There is some debate on whether it is safety climate or safety culture that is being measured by questionnaires (e.g. Halligan et al. BMJ Qual Saf 2011; Guldenmund Risk Analysis 2011). I would advise to refer to the debate and clarify what exactly the term should be in the context of this study.

Introduction:

3. Safety culture is widely accepted to have an impact on patient safety though the evidence is still quite small. Please refer to more current literature that actually show an empirical link between the both (e.g. Mardon et al. J Pat Safety 2010; Haynes et al. BMJ Qual Saf 2011). Please be more precise in the abstract: “A supportive patient safety culture is supposed to be an important.....”

4. The Medical Office Survey on Patient Safety Culture of the AHRQ is available since 2007 or 2008. Please add the information (my guess) that you were already developing SCOPE from the dutch HSOPS when the other questionnaire had been made public. I recommend a discussion of both contents and factor structure later in the article.

Methods:

1. Did you undertake cognitive testing? Did the pre-test panel meet personally? Was adaptation a team process with face-to-face meetings? Please add the
information in the manuscript.

2. Data collection and participants: As I understand the way of data collection the questionnaire had been "administered" online (Did participants have to log on to the web site, did they need a personal login or a practice login? Please provide this information in the manuscript). Do you know how many of eligible individual participants (in the 72 participating practices) actually took part in the study?

3. Data collection: Why was data collection stopped after five months (Had been a required number of participants achieved?)? Please add the required information.

4. Factor analysis: "Items that loaded on more than one factor were excluded". Due to my experience I would expect that at least some of the items load on more than on one factor with loadings of more than 0.4 (all the more there are few considerable inter-factor correlations). Please comment on this remark.

5. Internal consistency: I would like to indicate that there are different perceptions of Cronbach’s alpha that consider a value of 0.7 as acceptable (“recommended”) and of 0.8 as good (e. g. Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. J Pers Assessment 2003;80:99-103). On the other hand, cronbach’s alpha depends on the number of item per factor and therefore a three item factor with an alpha of 0.65 seems to be quite good. Please comment on this in the manuscript.

6. Please give the information on the way how you interpreted the average inter-item-correlation and evidence on the limit of 0.2 to 0.7 (there are recommendations that it should range between 0.2 and 0.4).

7. Construct validity: A limit of 0.7 is given to distinguish between high correlations and acceptable correlations. Please provide a reference since according to my experience a correlation of e. g. 0.6 is widely accepted as being quite high.

Results:

1. Please provide a response rate (either on practice level or individual level).

2. What was the limit for representativeness? By looking at the data it seems to me that the study population had more work experience than the national average. Had been statistical analysis been done or is this purely descriptive? Please comment on that.

3. How do you interpret the results of average inter-item correlations and item-rest correlations? Please add the interpretation in the manuscript.

4. The lower range of Cronbach’s alpha values: Please see my comment to the method section, no 5.

5. Did the 72 practices dispose of local reporting systems? A short information on the opportunities of participants to report would add valuable information to this result. Did you evaluate these opportunities as well?

Discussion:

1. I would like to recommend to revise the summary to “…the internal
consistency...were mostly satisfactory”.

2. In the paragraph on incident reporting the term safety culture is used normatively (how a culture of safety should look like). I would like to recommend to distinguish thoroughly between an analytic approach and the use of terms (that uses questionnaires to measure the various aspects of safety climate) and a pragmatic approach (see Guldenmund Risk Analysis 2011). The way you used the questionnaire safety culture could be “good” or “bad” and we believe that a “good” safety culture disposes of successful incident reporting and learning.

References:

1. Please revise the whole section carefully – there seem to be still some data on references missing (e.g. 25 and 29), some copied data from a reference manager (25 and 32; “Ref Type: generic”). Ref no. 14 is published 1999. And ref no. 9 is already published.

Discretionary Revisions

Results:

1. Construct validity: A correlation of 0.24 between scale 5 and scale 4 is weak compared to the other inter-scale correlations but not in terms of overlap of concepts in between the concept of safety climate or safety culture. On the other hand, there are correlations of 0.5 and higher suggesting that these factors might measure concepts with some possibly overlap. However, these values are not too high.

Discussion:

2. Differences between the hospital SHOPS and the general practice SCOPE: Maybe it is not different concepts but different appearances of the same concepts. That makes comparison difficult but not irrelevant. At the contrary, to my mind it is very interesting what makes safety culture in different contexts and therefore some comparison would be essential. Maybe, questionnaires are not the right instruments for these comparisons.

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

Quality of written English: Acceptable

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

Declaration of competing interests:

I declare that I have no competing interests’ below. If your reply is yes to any, please give details below.