Reviewer's report

Title: Evaluating cutpoints for the Mental Health Inventory (MHI-5) and the Mental Health Component Summary score (MCS) General Health Questionnaire (GHQ-12): A comparison of five different methods

Version: 1 Date: 26 July 2007

Reviewer: Hans-Jürgen Rumpf

Reviewer's report:

Kelly et al.: Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: A comparison of five different methods

The authors present data on two screening instruments for mental health, the MHI-5 and the MCS. By using the GHQ-12 as gold standard, they examine different cut-off points on grounds of different methods. The manuscript is clearly written and provides interesting data. The information on different methods to define cutpoints is very instructive. A clear strength of the paper is the large sample size of 14,669 individuals from the general population. The following aspects might add to the improvement of the paper:

1. A weakness of the analyses is the use of a screening questionnaire as gold standard to define cut-off values. The authors discuss this point briefly in the discussion section. I believe that this point should be stressed a little bit more to put data into perspective. A screening instrument as reference is always a makeshift and it is good to be aware of the inaccuracy of this approach. For example, assuming that the MHI-5 would detect all individuals with a mental disorder (sensitivity = 1.0) using the GHQ-12 as gold standard would reduce the sensitivity down to the detection rate of this instrument (e.g. .80). Depending on the extent to which the detected groups of both questionnaires overlap, misinterpretation of the real sensitivity and specificity varies. For example, assuming a sensitivity and specificity of .80 each for the GHQ-12 and the MHI-5 would result in a sensitivity and specificity for the MHI-5 of 1.0 in case of complete overlapping and of .60 if both questionnaires detect different "real" cases.

In addition, the gold standard has also an impact on defining cut-offs. Especially, a high number of false positives in the GHQ-12 would lead to higher cut-offs for the MHI-5. This might be the reason why studies using standardized diagnostic interviews have found lower cut-offs for the MHI-5. The discussion section should be adjusted accordingly. It should clearly be stated that the rather high cut-off found in this study differs from studies using diagnostic interviews and is likely to be overestimated.

2. On page 11, the authors state that in the field of mental health, no gold standard exists. It is true that all instruments are not perfect, however, there is agreement that standardized interviews like CIDI or SCID yield high validity and
reliability and should be preferred compared to screening questionnaires like the GHQ-12.

3. Related to that, the M-CIDI is regarded as a screener (page 10), however it is a standardized comprehensive diagnostic interview.

4. In the conclusion of the abstract the authors claim that "the Youden Index and (0,1) methods are most suitable for determining a cutpoint for the MHI-5, since they are least dependent on population prevalence". This is something that is independent from the data presented here. If generalizable cut-offs are aimed at, the methods used should not depend on different prevalence rates in different populations. Therefore, it is not fully clear why the authors used all the methods when they knew in advance that some of them would not fit the purpose.

5. In the same section, it is stated that "the MHI-5 performs remarkably well against the longer MCS". This is an interesting finding and should be put more in the focus of the paper. This should also appear in the results section of the abstract. As stated above, the analyses on cutpoints clearly suffer from methodological flaws. Therefore, the paper would profit from shifting the main focus.

6. Although details of the BHPS are published elsewhere, some important basic characteristics like the response rate should be given in the methods.

7. There should be a clearer distinction between results and discussion section. A lot of results are presented for the first time in the discussion and could be better presented in the results.