Author’s response to reviews

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

Authors:

Mark J Kelly (kellymj1@cf.ac.uk)
Frank D Dunstan (dunstanfd@cf.ac.uk)
Keith Lloyd (k.r.lloyd@swansea.ac.uk)
David L Fone (foned@cf.ac.uk)

Version: 2 Date: 26 October 2007

Author’s response to reviews: see over
Response to reviewers

Firstly, we would like to thank the reviewers for their comments and suggestions. In general, the thrust of the paper was appreciated and the accessible presentation of the material was acknowledged.

We now go through the reviewers’ comments in detail, beginning with referee 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.”

This is an important point and it was acknowledged in the original manuscript on pages 3 and 11. It has been further stressed in the revised manuscript on pages 3 and 11.

“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.”

Further detail regarding the impact of the GHQ-12 not being a gold standard has been included in the discussion section on page 11.

“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.”

It is possible that the higher cutpoints found in our paper compared with the literature may be a result of false positives in the GHQ. This point has now been acknowledged in the text of the manuscript on page 11. We still feel however, that a simpler and more likely explanation for the higher cutpoints is that other studies from the literature investigated more severe mental disorders than the common mental disorders.
“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.”

We thank the reviewers for pointing this out. It was acknowledged in the original manuscript that a clinically administered interview schedule would be preferable to the GHQ-12, however this point may not have been made sufficiently clear in the original version, and has now been clarified.

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

This comment refers to a section where the M-CIDI is incorrectly referred to as a screening tool. We apologise for this typographical error that has now been corrected.

“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.”

The referee is correct when he states that the conclusion that the Youden Index and (0,1) methods are the least dependent on prevalence is independent of the data presented here. However, the data are being used to investigate how dependent these methods are on the prevalence relative to one another. Are the differences negligible, or are they important? Figure 4 investigates this question. Moreover, these methods of choosing a cutpoint are some of the most popular amongst researchers and so it is of interest to compare them regardless of how suitable they may at first appear.

“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.”

We agree with referee 1’s assertion that this result is interesting, and it has now been included in the results section under its own heading on page 7. We have shifted the focus of the paper and we have included a discussion of this finding on page 9. The correlation between the scales has also been reported in the revised abstract on page 1.
“Although details of the BHPS are published elsewhere, some important basic characteristics like the response rate should be given in the methods.”

The BHPS is a complex study and in the original manuscript we preferred to keep our description brief, referring the reader to the Taylor reference for further details (578 pages long). However, in response to the review we have now expanded this section to include extra information on the study setting and response rate.

“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.”

We are grateful for this comment. Results are no longer presented in the discussion section.

Now we will respond to each comment of referee 2.

“Though this study has a number of merits, my main concerns are methodological. I have outlined these points below:

Major Points

a) My major concern is the use of the GHQ-12 as gold standard. The GHQ-12 is a screening instrument itself. Therefore, the authors are comparing two screening instruments rather than screening instrument vs gold standard. In my opinion, it is not appropriate to use the GHQ-12 as a gold standard. This is a serious limitation of the study. I would suggest not to focus on cutoff points but to provide information regarding the association between the instruments.”

Referee 1 also raised this concern and this point has now been further discussed in the revised paper on pages 3 and 11, as per referee 1’s recommendation. Furthermore, we have shifted the focus of the paper to include a more detailed discussion of the association between the two instruments. We have fully acknowledged that a limitation of the study is that the GHQ is not a gold standard, however we feel that the strengths of the study (large sample size, comprehensive selection of cutpoint choosing methods, participants being representative of the general British population, accessible presentation) greatly outweigh this weakness. Our results will also inform the second stage of our research programme in which we intend to undertake a further study using a standardised interview schedule in order to obtain a cutpoint for the MHI-5, as recommended in the conclusion of the paper.

“Minor points

Dichotomization of the GHQ-12 results in a lost of information. It might be more appropriate to compare several cutoff points for the GHQ-12 or use the Stratum Specific Likelihood Ratio approach as discussed by Furukawa (Furukawa TA, Goldberg DP, Rabe-Hesketh S, et al. Stratum-specific
The cutpoints for the MHI-5 and MCS were investigated in the original paper using different thresholds on the GHQ-12. This was illustrated graphically in figure 4 where the effect of varying prevalence was investigated by varying the cutpoint on the GHQ-12 between 1 and 12.

We thank referee 2 for bringing Furukawa’s paper to our attention. This paper offers stratum-specific likelihood ratios (SSLRs) as an alternative to cutpoints. The SSLR is the ratio of the probability of a given test result when the disease is present to the probability of the same test result when the disease is absent. These are more useful for diagnosis since they are based on the actual score, rather than whether or not it is below a cutpoint, and the SSLR provides a measure of the evidence from the test score. For a continuous measure, like the MCS, they rely on fitting an appropriate density function to the data and so are less easy to use. For public health purposes we believe that despite the simplicity of dichotomisation, cutpoints are a useful and intuitive tool. This is echoed by Hoeymans et al (2004) who say “A cut-off point will improve the usefulness of the MHI-5 to a great extent, for example for screening mental health problems in primary care or for calculating a health expectancy based on mental health”. The SSLR approach has now been discussed in the discussion section.

“It would be helpful if the authors could provide some information regarding the association between the instruments using a dimensional approach (e.g., correlation, structural equation model)”

Correlations between the scales were provided in the discussion section in the previous manuscript. They have now been moved to the results section, where they are more prominent.

“Did the authors control for age and gender, that means was there a similar association for males and females and younger/older subjects?”

The GHQ-12 does not have gender- or age-specific cutpoints to identify CMD and as such it would not be appropriate to use it to derive such specialised cutpoints for the MHI-5. The objective of the paper is to define a general cutpoint on the MHI-5 similar to the general cutpoint of three or more on the GHQ-12.

“It would be helpful if the authors could provide the wording of the questionnaires in order to identify overlap.”

We agree that it would be helpful to include the wording of the questionnaires in the paper. Unfortunately, we do not have permission from the copyright holders to do this and so we provide the reader with references where they can read the full questionnaires.
“There are different ways to compute a summary score (e.g. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res 2001; 10:395-404). Did the authors use the approach suggested by Ware et al.?”

The approach suggested by the developers of the scale (Ware et al) was used, as indicated in the original manuscript on page 3.

Finally, Competing interests and Acknowledgement sections have been incorporated into the revised manuscript.

We are grateful to the referees for their helpful comments. We trust that they will find our responses acceptable.

Mark Kelly
26/10/07