Reviewer’s report

Title: The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants

Version: 2 Date: 13 November 2012

Reviewer: Genevieve Gariepy

Reviewer’s report:

The manuscript reads much better, with notable improvements in the introduction and the discussion. Some additional comments and suggestions are provided below.

Major Compulsory Revisions

1. The evidence for external validity of the instruments is not clear. External validity is “the ability to produce accurate predictions among patients not included in the development of the system but from the same population” (Justice et al, Assessing the generalizability of prognostic information, Annals of Internal Medicine, 1999). Although the prevalence of mental disorders in the sample is similar to those of the target population, this does not show external validity of the instruments per say. For instance, the study sample could have less severe symptoms than the target population. The distribution of other risk factors could also vary between the 2 groups, such as age and sex distribution. Methods to check for external validity include internal validation tests, such as bootstrapping, and external validation tests, such as testing cut-offs on different samples of the same target population. One suggestion is to drop external validity from the objective of the study. Information on the comparability of prevalence of mental disorders in the sample vs target population could be provided in the Methods section, under Setting and Procedure. This would reassure readers on the representativeness of the sample.

2. In the discussion, the authors state that “both the K10 and the K6 outperform the GHQ-12 as to validity”. There is also a use strong language to highlight the inferior performance of the GHQ-12 such as “The GHQ-12 may not be suited for screening a population of long-term disabled persons [...]” and “Since the psychometric properties of the GHQ-12 are clearly inferior to those of the K10 and the K6, [...]]”. However, it is not clear from results if the AUC of the GHQ-12 is statistically significantly lower than the K10 and the K6, as confidence intervals overlap. And table 3 shows that the GHQ-12 cut-off may have better specificity than the 2 other instruments. The authors might consider conducting significance testing comparing the AUC of their instruments. Alternatively, the authors could use a more conservative language when comparing the performance of the instruments.

3. It is interesting that the threshold cut-off for the K10 is higher in the study
sample than in samples from the general population and primary care. However, the explanations for this are not obvious. First, it is not clear why the presence of psychological factors in long-term disability explains the higher cut-offs. Please add 1 or 2 sentences to explain. Second, it is not clear how prevalence differences in outcome influences cut-offs. Cut-offs are determined from sensitivity and specificity which are insensitive to outcome prevalence (compared to NPV and PPV which are sensitive to outcome prevalence). An additional alternative explanation might be that disability claimants score higher on the K10 because of concurrent physical symptoms (e.g., fatigue, lack of energy) that are items on the K10.

Minor Essential Revisions

4. p.11, “However, selection bias is not likely, since we found no significant difference as to the prevalence of most frequent mental disorders [...]”. As discussed in point 1, similar prevalence of mental disorders does not guarantee generalizability, although it adds evidence for it. Authors could use “less likely” instead of “not likely”.

5. Table 1, the column heading should be n(%) and the table should include the number of subjects (n) in each category

6. Table 2, the title should read “... and in the total population of disability claimants”.

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 declare that I have no competing interests.