Reviewer’s report

Title: Psychiatric Rating Scales In Urdu: A Systematic Review

Version: 1 Date: 13 August 2007

Reviewer: Kamaldeep Bhui

Reviewer’s report:

This is a potentially interesting and important paper. However, the restriction to instruments with validity data and no mention of those found with no validation does a disservice to the intentions of the study and the practical value of it. There is a lack of data given on validity coefficients, including area under ROC curves, for different gold standards, and what of different gold standards used? This should be taken into account in the review. This is an omission.

The schema for rating quality is not easily discernible.

Crosscultural validation usually means validation against an indigenously developed instrument by identifying appropriate local items, even if they fall outside of the conventions of psychiatric nosology; and not just translation and back translation; the authors don't make 'etic' and 'emic' distinctions that are important if the cultural issues are to be taken seriously. Although the authors argue for use of etic instruments, in the absence of emic data it becomes difficult to be sure of the treatment relevance of conditions, for example, what would we make of 65% of the population in some regions of Pakistan meeting criteria for depression as found in some etic studies? Will they benefit from antidepressants? There are many conceptual issues that could have been tackled but have not been.

The authors are reporting validations within a closed system of thinking, which is fine, as long as they acknowledge this. Are there any indigenously developed instruments? Now reporting these would have been valuable. Indeed, the problems in translation and validation, and debriefs from pilot work, can give clues to appropriateness of the adapted instrument.

Stretching to include comments on other inventories in a non-systematic way is unhelpful as it distracts from the main objective, and it could have been done much better.

I think the authors should consider all instruments developed on Pakistani populations the world over, for different conditions; by doing so they would provide a great service to the many researchers and clinicians who see Pakistani patients.

The paper will help researchers and clinicians choose an appropriate instrument in the Urdu language, but not for other populations. This is useful, but the details on extraction, synthesis, and the conceptual limitations of the more popular
instruments requires more attention.

The relevance is for one population, so the methods used to identify papers and extract data could be the innovation where other studies prioritising validation around a specific language could follow these methods if given in more detail.

The country of development, and the detailed translation, back translation, and gold standard data are important. The findings suggest a number of useful instruments; what they measure is quite another matter and whether this warrants treatment could be the focus of future studies.

The suggestions I am making will be quite a lot of work, but the paper would be much stronger and more useful to clinicians and researchers.