Reviewer’s report

Title: Exploring Differential Item Functioning in the SF-36 by demographic, clinical, psychological and social factors in an osteoarthritis population.

Version: 2 Date: 1 August 2013

Reviewer: M John G Bankart

Reviewer’s report:

Major compulsory revisions

a1. Statistical analysis section. I agree with the logic regarding assessing unidimensionality using large differences between the first and second factors, and small differences between the remaining factors, but no specific rule / method was mentioned. The authors state: “Unidimensionality was supported if there were large difference in eigenvalues between factor 1 to 2 and small difference in eigenvalues between 2 and 3”. The authors should quantify precisely what they mean by ‘large’ and ‘small’. Was there a formal rule for establishing the differences in proportion of variance between factor 1 and factor 2, and between factor 2 and factor 3, which were necessary in order to conclude that there was in fact a state of ‘unidimensionality’ pertaining to a particular factor? If so, the authors should state what this rule was.

a2. Effect of covariates sub-section. As this is an observational study, you cannot assume that covariates other than age were not different between demographic groups such as males and females. Was it possible to check this? You may have adjusted for age, but any DIF effects found could be due to imbalances in other covariates between groups. You should state that you have not ruled out this possibility.

a3. Conclusion section, 1st sentence. The authors suggest that they have uncovered DIF items. However, they may not have truly established the presence of DIFs. Not only have they not demonstrated that they have made appropriate adjustments for confounding variables, they have also not demonstrated that between any of the groups they are comparing, there really exist different underlying response probabilities. The authors should acknowledge this fact.

Minor essential revisions

b1. Background, paragraph 3. The authors give an example of DIF (measurement bias) in order to illustrate the concept. However, the association between depression and crying may well be different for men and women, as it is traditionally considered to be more socially acceptable for women to cry, and hormones are different in the two groups as well. On this basis, women probably do cry more than men, while controlling for emotional state, although this will be culture dependent to an extent. An item displays DIF only if people from different
groups with the same underlying true ability / tendency have a different probability of giving a certain response. With respect to the example given, the two groups probably have different mean underlying tendencies with respect to crying when depressed. The authors should give a different example, where it is less likely that there are real differences between the groups.

b2. Grouping factors section. The contrast between BMI categories, comparing \(<30 \text{ v } >30\) is currently labelled as normal / overweight v obese. This should be re-labelled as: underweight / normal / overweight v obese. Also, it should be \(<30 \text{ v } #30\) (or 30+).

b3. Grouping factors section. The authors state that they used bivariate median splits to dichotomise continuous and ordinal variables such as age. This is potentially problematic, as it can lead to an increase in either type I or type II errors, as well as being associated with other problems, such as inability to assess linearity / departures from linearity (Royston et al, 2006, Stats in Medicine). The authors should acknowledge the weaknesses of using such artificially / statistically derived dichotomisations in creating their categorical variables.

b4. Statistical analysis section. Testing assumptions. Why was principal components analysis rather than common factor analysis carried out (why was it deemed appropriate to use all of the variance rather than just the common variance ?) The reasons for this should be made explicit.

b5. Ordinal logistic regression model section. The authors should explicitly state what the dependent variable was.

b6. Measures section. Can you state which version of the SF-36 you considered in this study ?

b7. Related to the above point, the authors have not mentioned the situation where sub-groups report the same level of response, but the underlying response levels are in fact different. A major issue concerns how to establish that there are in fact real differences (or not) between sub-groups. This is not discussed in any detail by the authors.

b8. Table 1 (participant characteristics) : this table contains proportions and other information. The proportions should have the associated number in brackets after the %. The authors need to make it clear what the other information means. So for example, for Age (years), is this mean age (SD) ? BMI is given to 2 decimal places, whereas information for other numeric variables is given to 1 decimal place.

b9. Table 2. The abbreviations ‘var’ and ‘eigenval’ should be written out in full.

Discretionary revisions

c1. The authors might mention that selecting an optimal number of factors (in factor analysis) can also be done using parallel analysis (Horn, 1965).
method has gained in popularity in recent years.

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

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.