Reviewer's report

Title: The validity of aggregate income and education measures to proxy self-report: Does it depend on the patient group?

Version: 1 Date: 8 November 2010

Reviewer: Uzor C. Ogbu

Reviewer's report:

The authors have compared the agreement between individual-level and aggregated level data about income and education level from two sources. The topic is interesting and addresses a theoretical topic relevant to many studies. However, their approach raises a number of questions.

Major Compulsory Revisions

1. In the early part of the article, the authors appear to portray the comparison as one of self-reported information versus aggregated information obtained from alternative sources. However, in the limitations section they state that the aggregated information used is based on self-reported census data. A change in terminology from self-reported, to individual-level and aggregate level seems to be more appropriate. They should also include the description of how the census data was obtained in the methods section.

2. Did the authors account for respondents currently enrolled in a university when classifying the level of education attained?

3. Did the authors consider other categorizations of education? For example, grade school, high school, and university.

4. The authors should include information in the text about the sampling methods or representativeness of the surveys used, as this is essential to the interpretation of their findings.

5. Can the authors comment on the generalizability of their findings as it relates to countries/surveys that derive their aggregated data from tax returns?

6. The authors conclude that there may be differences in patient groups using RA patients as an example but then relate this difference to age. Might they not find similar differences in the DM group had the mean age been higher? Is this then a difference in self-reported income across patient groups or a methodological flaw in the survey related to age?

7. Can the authors comment on why the agreement between individual- and aggregate-level reports varies by diagnosis? Is it possible that respondents were less truthful either in the surveys or in the census?

8. What do the authors recommend? The authors point out that often in
administrative data, individual level information is not available. Should aggregate-level census data not be used in these situations? Does it create a bias and if so which direction do they expect it to be? Are there statistical models that might take into account the aggregated nature of the data?

Minor Essential Revisions

9. The title of the manuscript does not give a clear description of its contents. It gives the impression that it focuses on a stratified analysis. The research question in the abstract indicates that the purpose is to “…determine the level of agreement between aggregate-level and self-reported measures of income and education among three distinct patient groups…”

10. In the description of the analysis in the abstract, the authors appear to focus on comparing agreement between the individual-level and aggregate level reported income, and the differences in population groups for the reported education level.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.