Reviewer’s report

Title: Socioeconomic status and race/ethnicity independently predict health decline among older diabetics

Version: 1 Date: 9 June 2011

Reviewer: Frans Tan

Reviewer’s report:

Review ‘Socioeconomic status and race/ethnicity independently predict health decline among older diabetics.’

By Emily Joy Nicklett

The paper describes a study that examines the effect of race/ethnicity and socioeconomic status on health decline over time among older adults with diabetes. The author claims that it is the first study that examines this relationship based on a sample of older diabetics. I have no reason to doubt this since I am not an expert on this topic.

Some remarks regarding the problem formulation.

The paper is well written, but the problem is vague and at least not well formulated.

Background, fifth paragraph, the author states that ‘... are analyzed independently and in conjunction...’, and in paragraph six ‘The primary aim is to examine how socioeconomic characteristics and race/ethnicity- both independently and in concert- predict...’. The problem here is the meaning of ‘independently and in concert’.

A mistake that is often made by researchers is the interpretation of an ‘independent effect of a predictor’ and is often confused with ‘an effect irrespective of that of a third variables’. The first one should be analyzed by model 3 and the second one is meaningless since it may be a spurious relationship.

To explain this and simplify the situation, suppose one is interested in the independent effect of socioeconomic status on health. Suppose further that socioeconomic status (SEC: high/low), race/ethnicity (Race: whites/blacks) and health (Health: good/bad) are all dichotomous. If, for example, there are more blacks with low than high SEC and more whites with high than low SEC, then SEC and Race are clearly correlated. If further, whites are healthier than blacks, then Race is a potential confounder in the relationship between SEC and Health. The author analyzed the relationship between SEC and Health (model 2) excluding Race. So if the author would find that high SEC is associated with good Health, then this association could very well be due to the fact that there are more whites with high SEC in the sample. Consequently, the relationship between SEC and Health and excluding Race (model 2) still depends on the
distribution of Race in each level of SEC (and may differ from sample to sample) and hence you cannot consider the relationship (which may be spurious) as an independent effect of SEC on Health just by excluding Race from the analysis. Model 1 and Model 2 are not only superfluous, it is also incorrect to perform such an analysis simply because of the confounding effect. Model 3 is the method to analyze the effect of each of the predictors on Health. This makes it possible to analyze the effect of, for example, SEC among all people of the same race.

The effect ‘in conjunction’ and ‘in concert’ is not clear at all to me. It may however be interesting to analyze interaction effects (or perhaps this is what the author means). The author did not consider this.

Some remarks regarding the sample and analysis strategy:

Analytic sample, first paragraph
Data are drawn from the RAND (2008) combined data files and imputations. The author should give more information about this. For example, which type of imputations were made (the covariates or the outcome or both and single or multiple).

Covariates Analyzed: Height is carried forward to impute the missing cases. Why do you want to do this? Is it time dependent? Why not consider height as a constant? (for example, average of all measured heights of the person).

Analytic strategy, second paragraph. The author considered the multilevel cumulative logit regression models. As such this is a good choice due to the ordinal character of the outcome. However, the reason mentioned by the author is not correct (… which considers… assume continuous variables). The reason of the favored ordinal logit analysis is because of the specific distribution of the responses (Using OLS will lead to biased estimates in general and using normal models will invalidate the distribution of the test statistics).

Did the author use GLAMM option in Stata and perform numerical integration and if so what are the random effects?

Third paragraph. All analyses were linear and were conducted… What does the author mean by this? Logit models belong to the class of generalized linear models and are certainly not linear.

Some remarks regarding the Results section.
I miss plots of observed and estimated probabilities, so that we can check what the model is describing.

The problem of analyzing the data using this technique is that the distribution of the loglikelihood function is not known. Consequently statistical testing of the estimated regression parameters cannot be trusted. How does the author deal with this problem?
I assume that the author makes use of the proportional odds assumption. It should be mentioned in the paper and some justification of this assumption should be made.

In Table 1 and 2 the author presented ‘Mean’ values (third column). In fact these are percentages, except for the last four rows.

The self rated health is presented as column percentages. I think it is better to present row percentages.

I do not understand what the numbers of the last four rows (Self rated health) stand for.

Minor remarks
Repetition of content: Abstract, results section, ‘controlling…remain’ and ‘Including…Hispanics’ do have the same meaning. Perhaps add ‘…differences remain, i.e. including socioeconomic…’. Or rephrase the sentence.

Background, fourth paragraph. Longitudinal research has also indicated that racial/ethnic...

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

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**
I declare that I have no competing interests