Reviewer's report

Title: Estimation Methods with Ordered Covariate Subject to Measurement Error and Missingness in Semi-Ecological Design

Version: 1 Date: 20 April 2012

Reviewer: Kofi Adragni

Reviewer's report:

I congratulate the authors for this document. They present the case that, a constrained group-based strategy in estimating parameters involved in a semi-ecological design where ordered covariates are observed with measurement errors and possibly with missing values has desirable bias-reducing properties and is easier to implement than a constrained expectation-maximization method.

My review of this paper comes from two perspectives. First, I am a trained statistician who is familiar with the statistical methods employed in this article. Second, I am not an expert in semi-ecological design and cannot well evaluate how novel this work is and how it builds on past research. Consequently, my comments and suggestions are guided from the first perspective.

Overall the paper is well presented and the authors introduced their statistical method in a somewhat balanced manner. The research question was well formulated. The authors used a simulation study to compare four approaches: (1) naive complete-case analysis, (2) group-based strategy, (3) constrained group-based strategy, and (4) constrained expectation-maximization. They applied these methods to a relevant dataset.

It must be stated that this is a simulation study and finding from this study are rather empirical. There seem to be a lack of references (bibliography) to existing methodologies. Is this the first publication on semi-ecological design study using EM algorithm? What methods have been used in the past? Why is this work relevant?

The title and abstract convey what has been found and the writing is acceptable, but I have some criticisms:

Major Compulsory Revisions

1. On page 7, before section “EM with Measurements…”, you suggest that if there is no available information on sigma^2_y, one may use AIC or SBC to select it. How can you, since you have no candidates and the range of sigma^2_y is infinite? The suggestion does not seem correct.

2. The introduction of the isotonic regression in the Methods section seems abrupt. Its purpose was not obvious until the last two sentences on page 5. Perhaps the authors should mention that CGBS uses the isotonic regression and continue to present what it does.
3. On page 7, in section “EM with Measurements…”, the sentence “Observations containing the cases with missing values in covariate can also be accommodated by the constrained EM algorithm if covariate are ordered” suggests that you are the first to consider this setup. The use of ordered covariate does not seem to show anywhere in your derivations. Please address this and/or provide references.

4. Line 9 from bottom of page 5: No information was provided on how Table 1 was obtained and what it provides up to this point. The authors can either elucidate how Table 1 was obtained or make their point in the simulation.

Minor Essential Revisions

5. On the third line from the bottom of page 4, what is Z?

6. Please provide a reference for PAVA.

7. The authors should fix the following sentences: Page 2, line 5: “Naïve and CBS…”; page 3 line 5: “… incurs…”; page 3 line 7: “… to assume groups are …”; page 7 line 1 first sentence in section with “Wu summaries…”; page 10 line 1 in new section “… measures…” should that be measurements?

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

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.