Reviewer's report

Title: Dealing with Missing Data in a Multi-question Depression Scale: A Comparison of Imputation Methods

Version: 1 Date: 11 September 2006
Reviewer: Kosuke Imai

Reviewer's report:

General
1. Is the question posed by the authors new and well defined?
   There have been many (theoretical and empirical) studies examining various methods to deal with missing data. And hence, in that sense, what the authors are trying to do in their manuscript is not new. However, their contribution is to address this important question in a particular context; i.e., with respect to multi-question depression scale.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?
   See my comments below.

3. Are the data sound and well controlled?
   The data presented in the manuscript are interesting and a good example.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   I think so, though in some places how the authors used the data is unclear.

5. Are the discussion and conclusions well balanced and adequately supported by the data?
   See my comments below.

6. Do the title and abstract accurately convey what has been found?
   I think that the title is appropriate. The abstract needs some work.

7. Is the writing acceptable?
   I think the writing is acceptable.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. When the authors say “data were deleted randomly” (see for example the second para of page 4), they actually mean that the data are missing completely at random, which is the condition known as MCAR in the statistics literature (see e.g., Little and Rubin, 2002). It is important to keep these technical words as accurate as possible to avoid possible confusions.

2. When the authors say “a non-random simulation” (the first sentence of page 5), their procedure is basically missing at random (MAR). In the statistical literature, there is another important condition called “non-ignorability." The authors should consider conducting a simulation study for this condition as well.

3. Overall, there should be more discussions about how theoretical results connect the authors’ empirical findings. For example, we know from Little and Rubin (2002) that under MAR the multiple imputation is supposed to work well. How does this theoretical result relate to the simulations the authors conduct? In other words, what theoretical issues do the authors’ simulation exercises shed light on?

4. The authors incorrectly describe the assumption of the MI method when they state “The underlying assumption of the MI method – that the missing data are missing completely at random – is not satisfied in these scenarios” (page 11). In fact, the assumption of the MI method is MAR. I think it is very important for the authors to clearly and correctly state what assumptions are required for each of the methods and why simulation studies they conduct are important for understanding theoretical issues regarding these assumptions.

5. 2nd paragraph of page 12. The main advantage of the MI method is that unlike single imputation methods the resulting confidence intervals have better coverage probabilities. Therefore, it is important for the authors to examine the uncertain estimates as well as the point estimates. The authors briefly mention this point on page 14, but I think this is “the main” advantage of the MI method.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. From the authors' description alone, I am not sure how “Single Regression” works.
2. Page 8. From the authors' description alone, it is not clear exactly how multiple imputation is conducted. For example, what is “a specified list of characteristics”? Do they include all questions?
3. 3rd paragraph of page 6. Are their missing values in age and gender?

Discretionary Revisions (which the author can choose to ignore)

It would be very useful if the author can provide some advice to applied researchers about what they should do when analyzing the data with missing values. The current manuscript does not say much about what applied researchers should do in practice. In the statistical literature, some recommends sensitivity analyses, and others recommend the method of bounds. I wonder if the author can go one step further and make some useful and clear suggestions.

References