Author’s response to reviews

Title: Dealing with missing data in the Center for Epidemiologic Studies Depression self-report scale: a study based on the French E3N cohort

Authors:

Noémie Resseguier (noemie.reseguier@yahoo.fr)
Hélène Verdoux (hverdoux@ch-perrens.fr)
Roch Giorgi (roch.giorgi@univ-amu.fr)
François Clavel-Chapelon (francoise.clavel@igr.fr)
Xavier Paoletti (xavier.paoletti@curie.net)

Version: 3 Date: 10 February 2013

Author’s response to reviews: see over
We are very thankful to the reviewer for his comments and suggestions.

Please find below point-by-point answers.

**Minor compulsory revisions**

1. OK, except
   a. the statement “in regression analyses of the association between an outcome and the CES-D scale, the imputation model can be safely adjusted on the CES-D items and on the covariates of interest” in the discussion is misleading:
      i. it’s not just “can safely”, it’s “must”
      ii. the imputation model must include the outcome as well; better to replace “the covariates of interest” with “all other variables in the analysis model”
      iii. minor: “adjusted for” not “adjusted on”

   We have modified these three points in the text.

   b. The abstract still doesn’t recognize this point: “The objective of this study was to investigate the best approach for handling missing data in the CES-D scale” – perhaps add “when interest lies in estimating the mean and prevalence”.

   Actually, multiple imputation appears to be a valid approach do deal with missing data, not only to estimate descriptive summary of the CES-D but also in a more general framework. Clearly, if the CES-D is to be used for regression, additional aspects must be considered; nevertheless the missingness mechanism should not be different when the CES-D is used in a regression analysis. In line with our conclusions, we think that missing values in the CES-D scale can be considered as ignorable, that multiple imputation can be used to perform analyses, and that it is worth investigating the robustness of any descriptive or etiologic analysis under scenarios for MNAR data.

7. Abstract, “missing data mechanism appears to be ignorable” – I still don’t think your results justify this statement. Rather, the qualitative study found nothing suggestive of non-ignorability, and your analyses suggested findings are reasonably robust to departures from MAR. I think this sentence should be modified.
We have modified this sentence as “does not appear to be nonignorable and estimates are robust to departures from ignorability”.

11. “These intervals did not include the point estimate based in complete data only, suggesting that MVs were not MCAR.”: I’m still not happy with this, as the intervals considered do not allow for ordinary statistical uncertainty due to chance (they only allow for ignorance due to not knowing the missing data mechanism). Essentially, you are saying that one estimated prevalence is greater than another, therefore there is a true difference, without conducting a significance test. Maybe it’s enough to append “although the differences are not statistically significant”.

We agree with the proposed limit in the interpretation of this part of our results. We have appended the proposed limit “Although the differences in prevalence were not statistically significant, these intervals did not include the point estimate based on complete data only, suggesting that MVs were not MCAR.”.

15. I think “response category” should replace “modality of interest” which still means nothing to me (change both in footnote and in the text on p11).

We have replaced “modality of interest” with “response category” in the suggested parts of the manuscript.

Discretionary revisions

2. The re-structuring is good, except that the “simulation study” section on p10 now refers to the person-mean approach which hasn’t yet been defined – perhaps move “simulation study” later?

We have moved the definition of the person-mean approach in the simulation study section.

Two new changes:

p6, “hormonal life” – change to “hormonal status” or “history of hormonal symptoms”?

We have replaced “hormonal life” with “history of hormonal phases”.

p9, “binomial negative regression” should be “negative binomial regression”
Done.