

Reviewer's report

Title: Multiple imputation for estimation of an occurrence rate in cohorts with attrition: a simulation study

Version: 1 **Date:** 28 June 2010

Reviewer: John H Kalbfleisch

Reviewer's report:

My aim of this review is to provide the authors with comments that can expand the eventual readership. This is an interesting article and it's message should be known to data-analysis persons.

MAJOR COMPULSORY REVISIONS

(1) Footnotes in Tables 2 and 3. First footnote to indicate that "coverage" reflects the % of 500 simulations-runs where the 95% Confidence Interval contained the true occurrence rate. The second footnote should explain the "Standard Error" columns. Is the "standard error" the variability measure of the value in the respective "Estimate" column, or the standard deviation of the 500 event rate estimates (clarify for readers)? In place of adding table footnotes, the Methods Section can elaborate on the content of the two suggested footnotes. see(2).

MINOR ESSENTIAL REVISIONS

(2) This is for the authors consideration. Comparison of simulation results for the two methods can also be graphically provided by histograms of P(E) estimates (one histogram for each method); a suggestion would be to select one of the scenarios in the current Table 2).

(3) Approximately line 35. Some readers will have difficulty understanding the "Multiple" in multiple imputation - an elaboration would benefit readers. ["multiple" implying more than one attrition case or "multiple" suggesting that several calculations are performed in the imputation process(multiple covariates – or calculations performed at multiple time points beyond the first time point)].

DISCRETIONARY REVISIONS

(4) I feel this manuscript will be better received by a larger readership if the authors provide an example of a clinical study that is MAR and explain why it is not MACR. Also, provide the Kaplan-Meier and Multiple Imputation P(E) estimates and standard deviation of the simulated estimates). Readers will see that the results of the two computational strategies differ, and therefore wonder which result would be preferable (hence, motivation for the current manuscript). Such an example could be added at the end of the Result Section.

(5) Table 2 and 3 could be combined by inserting an additional column to identify MAR or MCAR attrition type.

(6) Are there any situations where multiple imputation is inferior to K-M (appropriate for the discussion section)?

(7) What happens when in a study where some cases are MCAR and some are MAR – a mixture of attrition mechanisms? (it would seem multiple imputation would be advantageous, as in MAR, but not to the same extent).

(8) The simulation was devised so there are 4 time periods, baseline and a sequential set of 3 follow-up data capture points. Many research studies in which survival analysis is appropriate either measure individual survival times or have many more planned occasions of data capture (once a day, once a week, once a month or even once an hour as in some animal studies). Can the authors comment on simulation results for K-M vs Multiple Imputation for studies with a larger number of time intervals (say 50 periods instead of 4)?

(9) line 197, “particularity” to “characteristic”

(10) line 199, “works” to “publications”

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. John H. Kalbfleisch, Ph.D.