Reviewer’s report

Title: Assessing quality of life in a randomized clinical trial: Correcting for missing data

Version: 1 Date: 30 January 2009

Reviewer: Juni Palmgren

Reviewer’s report:

This is an exceptionally clear and well written paper on a methodological topic of considerable importance. The paper deals with assessing quality of life measurements in a panel study design subject to dropout - i.e. missing at random. Three methods are proposed (LI, IPW and MP) and used on a randomized study comparing two cancer treatments. Results from the different methods for treating missing values are presented and used on real data.

General comments on methods:

The methodological part is very well written, but it remains a bit unclear what type of reader is in mind: a statistician or a clinical - or both? A statistician would understand the concept of predictability on the middle of page 5, but a clinical would not. On the other hand a statistician knowing the world of regression is a bit surprised by the detailed description of indicator variables when describing a categorical predictor at the top of page 10. As a general comment: where possible it would be good to open up formulas and theoretical concepts in the methods section by adding a sentence or two in words.

You are dealing with an ‘immortal cohort’ and on page 4 say that you have not distinguished between missingness due to patients dying or dropping out for other reasons. This would require a bit more elaboration, in terms of what biases could occur by assigning QoL values to a person after death. It also remains unclear how the QoL scores (1-7) with an absorbing state D for death (page 9) is compatible with the immortal cohort approach?

You are treating the missingness as monotone in one type of analyses, and in another type of analysis you use single imputation for non-monotonous missing values before applying the monotone missingness strategy. This latter approach appears a very useful and sensible thing to do. A comment on why multiple imputations were not considered could be good (not needed, too tedious or something else..). I would also suggest using other terms than ‘valid’ and ‘non valid’ score (page 10 and many other places), since the latter are certainly valid when using the imputation strategy.

General comments on the application:

You choose to report results on two items of the 30 item QoL EORTC questionnaire, as an illustration of your methods. Since the analyses and the
reporting of results are essentially similar for the two items nothing is added by including both. I would suggest cutting the empirical part to one item only. This will cut the number of tables and figures substantially.

If we consider the item 29, I find Table 1 important. Figure 3 is also interesting. But I find Table 2 a bit obsolete. If you want to indicate that standard errors increase over time, perhaps a bar plot from one of the methods could be included at the bottom of Figure 3 with a phrasing in the text that the three methods gave very similar standard errors. Figure 4 is fine - but would you not want to include at least the mean score from the observations only. After all, this is the comparison of interest and the very point where the usefulness should show up of treating missing values properly. I suggest omitting Figure 5, but retaining Figure 6, with a comment on why you don't include MP, and perhaps now use Figure 4 as a comparison (reflecting what you gain from using the single imputation - precision I would guess?).

Tables 2-6 (standard errors presented in some other condensed fashion) and Figures 7-10 could be omitted altogether.

Discussion:

Please, give the reader some general hints on which of the three procedures you would suggest using in practice - and why? Do they pick up different types of missingness? Is one easier than the others to programme or to compute? Also, these days it is valuable to provide the reader with a source where to find code to mimic the analyses. What software have you used? Do you use standard routines, or is the code placed on a webpage?

Level of interest: An article of importance in its field

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.