Reviewer’s report

Title: Testing the treatment effect on competing causes of death in oncology clinical trials

Version: 2  Date: 28 January 2014

Reviewer: Ruwanthi Kolamunnage Dona

Reviewer’s report:

The paper discusses a simulation approach for comparing 3 available methods for modelling the competing risks. The addressed problem of misclassification of the course of death in a real data setting is itself interesting, and it is likely to have impact on the way these data were analysed in the past with similar data.

Major Compulsory Revisions

(1) In page 6, 2nd para in Results and Discussion and in page 9, 2nd para in Conclusions: Authors mentioned that empirical rejection probabilities are very close to 5%. But this is not always the case as of results in Table A.1. For example, CD, rho=-0.75, cens=0%, Pe rejection prob is 0.028, which is relatively below 5%. Similar situations for scenarios rho=-0.375, cens=25%, Gr; rho=0.375, cens=0%, Pe; and some also in NCD. In CD, rejection probabilities for some scenarios are relatively high; for example, rho=0, cens=25%, Cs it is 0.072; and in NCD rho=-0.75, cens=0%, Pe, it is 0.072 and rho=0, cens=0%, for all methods it is 8% rather than 5%. Please explain these changes from the 5%. Similar query goes for those relatively high and low probabilities in Table A.2.

(2) Page 8, first para, 2nd sentence from the end: In Pe method, which deaths would be reclassified as CD? It is mentioned that 77 unknown deaths and 26 NCDs, why these 26 and is this number correct?

(3) The main conclusion “Peto’s test as the most robust to misclassification of the primary event over the Gray’s and Log-rank (course-specific) tests” is not always true from the given simulation results. Therefore this conclusion should be mentioned cautiously by highlighting the specific scenarios (such as scenarios 3 and 4).

Minor Essential Revisions

(1) Figure 2 - please specify what is CT and Control in the title.

(2) Figure A.2 and A.3 should be drawn on a better vertical scale rather than 0 to 1 so the reader can see those empirical rejection probabilities from the 3 methods clearly with respect to the nominal 0.05.

(3) In page 6, last para in Results and Discussion: CS test: lowest alpha for CD as of Table A.3 is 0.03, not 0.02.
(4) Revise the title of Table A.3 to give what results you show for NCD. The rows for CD show rejection probabilities while those rows for NCD show power? Same revision goes for Tables A.4, 5, 6, 7, 8.

**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.