Reviewer’s report

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

Version: 2 Date: 12 February 2014

Reviewer: Geraldine Rauch

Reviewer’s report:

This is an interesting paper with a relevant topic. But I think, that the impact of the paper could be improved by considering and discussing the following aspects:

Major Compulsory Revisions:

• The authors may wish to consider


  as an additional reference as many of their issues are discussed in there as well.

• As stated by Pinitilie (Reference 11) and in the paper stated above, Gray’s test and the cause-specific hazard test assess different aspects of competing risks. In this sense it is not completely fair to compare their power properties because they aim to show different things. Gray’s method investigates the subdistributional hazards which can be used to obtain the cumulative incidence functions. By this the plot of the CIFS is directly related to Gray’s method. The CIF provides unbiased estimates of the event probabilities, however they may be difficult to be interpreted in case of existing competing events. The cause-specific hazards analysed via the standard logrank test can be used to test the effects of competing events independently. The related plot would be the Kaplan-Meier-Plot. However, in the presence of censoring the Kaplan-Meier-estimator no longer provides unbiased event probability estimators. This should be clearly stated in the paper. In the presence of competing risks, it is always better to plot the CIFs instead of the Kaplan-Meier-curves for graphical representation. So the cause-specific logrank test should be recommended, but the Kaplan-Meier-plot is not appropriate.

• Another major issue is the problem of correlated events. The cause-specific logrank test statistics between two disjoint events are always uncorrelated, see


  It would be more interesting to report the simulated correlation between the test
statistics instead the one between teh survival times, because ist the distribution of teststatistics that influences the power.

- 500 replications for simulation are far too less. Minimal number of runs should be 10’000 up to 100’000.

**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.