Reviewer's report

Title: Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models

Version: 1 Date: 24 February 2011

Reviewer: Yingwei Peng

Reviewer's report:

This paper considers an extension of the flexible parametric survival model for relative survival with a cure fraction. As the existing model, the proposed model employs cubic splines to fit the log cumulative baseline hazard in the non-mixture cure model for relative survival. To accommodate the cure patients, the proposed extension restricts the log cumulative hazard approximated by the cubic splines to be constant after certain time point. The proposed model is interesting and should be useful in population studies even though it is a simple extension of the existing model. However the paper lacks details of the estimation method for the proposed model and details of the data analysis.

Major Compulsory Revisions:

1. Page 6: Do you require that $s(x; \gamma_i)$ in (15) be zero after certain time point? It is not specified in the paper.

2. Some details of the estimation method should be given in the paper. Otherwise, it is difficult to follow some discussions. For example, what do you mean by "... orthogonalise the spline variables ..." on page 7?

3. Data section: What are the covariates considered in the model, age and calendar time? I suppose that there are other covariates for each patient from Finnish Cancer Registry. Comments on how the covariates are selected or determined should be useful here.

4. Page 8: It is not clear based on what evidence the authors believe that the flexible parametric cure model overestimates the cure rates in some cases. The paper does not report any simulation studies to investigate the performance of the proposed estimator. The author should consider simulation studies to support their conclusions in the real data analysis.

5. Page 9: It is not clear how the age and calendar time enter the model? Did authors group the data based on the values of age and calendar time and fit the model to each of the subsets? If yes, why can't the two covariates be treated as two continuous covariates and used as $z$ in (15)?

6. Page 9: If a cure fraction is believed for the colon cancer data, would comparing different cure models be more interesting than comparing a cure model with a non-cure model?
7. Page 10: I don't understand this statement "This shows how hard it is to rely on formal tests for assessing the assumption of cure, especially in large datasets where comparisons between models are often significant." If it is easy to become significant in comparisons between models in a large dataset, why can't cure be significant or the cure model be significantly better than a non-cure model if it does exist?

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.