Reviewer's report

Title: Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer

Version: 2 Date: 6 January 2010

Reviewer: Thomas Scheike

Reviewer's report:

First, I think the paper is improved and I agree with most of the changes. Obviously, however, we can have different views on how to do this in practice. I here give some last comments on this very nice and important paper that in my view should be accepted subject to some last final editing.

In my view the most useful thing to do is to use the cumulative residuals, that also seem to be quite powerful in detecting non-proportionality, and as I said in my previous report these cumulative residuals are implemented in SAS and the timereg/gof package in R. In my view the resulting plots of cumulative residuals are easy enough to interpret, and this is what we do in our paper Cortese G, Scheike TH, Martinussen T. Stat Methods Med Res. 2009 Jul 16. [Epub ahead of print] Flexible survival regression modelling.

where we also have a rather detailed discussion of the proportionality testing in practice. The cumulative shows the timevarying effect compared to an average effect, if the cumulative goes up first it is larger than the the average effect initially, and then it must go down at some other point !

The smoothed schoenfeld residuals are clearly related but the amount of smoothing is crucial and the standard errors of this is plot has at best the usual problems of a standard errors on a smooth. In this case it is worse, however, if one wants to interpret the standard errors as error bounds for the timevarying effect of the log-relative risk beta(t), because they are based solely on the cox model and not the extended model with beta(t). Also the idea of the smooth is that it should yield a rough estimate of beta(t) for a cox model on the form exp(beta(t) x), and this is correct in a certain sense. The smooth is a one-step Newton-Raphson estimator based on an initial constant estimator see Martinussen and Scheike (2006) for more on this. Clearly such an estimator will have some problems and will for example not be consistent unlike other estimators, such as the sieve estimator (partitioning of the time-axis) and the spline estimator of Zucker and Karr (1990).

I agree that the procedure tends to work well, but one should be aware of its shortcomings and the fact that the standard errors are also problematic.
The test reported as test of zero slope is in fact the test for whether \( \beta \cdot t \cdot x \) is significant, and therefore not related to the plot of the smoothed schoenfeld residuals, and there is nothing graphical about the test. This should be clarified in the manuscript.

I am sorry that I have misunderstood Table 4, but I was mislead by the wording where you suggest that you group according to the survival times! In reality you are simply suggesting to allow the Cox model different effects on different parts of the time-scale, this is also a standard technique that is very useful. So I would suggest that you change the wording of Table 4, you are not doing an analyses for event times shorter or longer than 4.3 years, but you are estimating the relative risk for different periods on the time-axis as you explain very nicely in the paper on the bottom of page 11. I am sure I overlooked this in the first version also (sorry!).

Finally, one last comment is that even though we all interpret the tests for non-proportionality one at a time the tests considered in this paper, however, in reality just give a test for whether or not cox's model is valid.

If we consider two correlated covariates and only one has severe non-proportionality then the tests considered in this paper typically reveal lacking proportionality for both covariates.