Reviewer's report

Title: Genetic analyses of smoking initiation, persistence, quantity, and age-at-onset of cigarette use in Brazilian families: Baependi Heart Study

Version: 1 Date: 7 September 2011

Reviewer: Terry M Therneau

Reviewer's report:

I found this overall to be well written. There are 3 serious questions about the survival analysis, and a couple of discretionary points.

1. The variable "age" is never specifically defined. Is it age at onset, age at enrollment to the study, or age at a fixed time point?
   This becomes particularly clear in looking at the survival analysis for age at onset.
   a. If "age" = age at initiation, then the analysis is invalid. You cannot have the same quantity on both sides of a relationship.
   b. If "age" = age at some fixed time point, then the variable is equivalent to adding birth year to the model, and is a test for calendar time effects, e.g., that smoking started at an earlier age in prior times than it does today. If we assume that age is in years, then the HR of 1.04 suggests that older people we much more likely to start smoking at a young age.
   c. If "age" = age at study enrollment, then the interpretation is a similar to b, assuming that the enrollment period is short compared to the age range. We would need this information to interpret the variables.

2. In the Cox model the variable "sex" was presumably deleted from the model when it was found to be "not significant". A fit that includes interactions without the main effects is, however, very sensitive to the variable coding. That is, using sex=0/1 versus sex=1/2 will give different answers. Please give details of how this was done.
   Interpretation of the results also requires knowledge of how the sex variable was coded. Does the HR of .98 mean females start out younger than males, or older?

3. "The individual risks of age ......and 65% higher in males than the overall average risk ....".
   I found this sentence and subsequent ones very confusing. The std for males is sqrt(.25)=.5, exp(.5)= 1.68, exp(-.5)=.60. This says that the "average" familial effect is for their members to have a relative log-hazard that is .5 higher or lower than the population as a whole.
   (60% less or 68% more on the multiplicative scale).
I don’t know where the last sentence comes from. Assuming additivity I get a variance for females if \(.31+.28, \sqrt{.59} = .768, \exp(.765) = 2.1\) or double the risk.

4. I’d be surprised if the effects of age or sex satisfied proportional hazards. Did you look at this?

5. Minor comment: I would have found column percents more useful than row percents in table 1.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare no competing interests.