Reviewer's report

Title: Estimation of age- and stage-specific Catalan breast cancer survival functions using US and Catalan survival data.

Version: 3 Date: 17 December 2008

Reviewer: Philip Rosenberg

Reviewer's report:

The authors have done a really nice job responding to my suggestions.

I have just one major inquiry about the revisions that I would like to see clarified. This pertains to the implementation of a 'parametric bootstrap' procedure to estimate the variance of the US hazard rates (page 5, paragraph 2 of the revised manuscript).

The authors correctly state that under the model, the number of breast cancer deaths in interval t follows a Poisson distribution with mean equal to the smoothed hazard rate from the model, times the observed numbers of woman-years in the interval. The language as written suggests that for each t the authors then generated 10,000 random Poisson counts from the Poisson distribution with the given mean, divided by women-years, and used the 2.5 and 97.5 percentiles of these hazard values to estimate a confidence interval.

If so, I do not believe these values would be correct, because the values do not reflect the uncertainty of the parameter estimates. Therefore, the authors should clarify, justify, or adapt the calculations of the standard errors.

In general, there are a number of ways to implement a bootstrap, but I believe a standard approach for this situation would proceed as follows.

1. Fit the Poisson regression model of the smoothing spline to the observed data, as done.
2. For each t, t = 1, ..., T, calculate a Pearson residual,
   \[ Z_t = \frac{O_t - E_t}{\sqrt{E_t}} \]
   where
   \( O_t \) is the observed number of events in interval t, and \( E_t \) the expected number under the model.
3. Repeat 10,000 times:
   3a take a random sample of size T from the residuals, \( Z_{tb} \), and use these values to create a random sample of count data generated under the model, using the formula
   \[ O_{tb} = \max(0, E_t + Z_{tb}\sqrt{E_t}) \]
   3b refit the smoothing spline model using \( O_{tb} \), t = 1, ..., T, and the observed women-years, to create a resampled hazard rate function generated under the
model.

4 From the 10,000 replicate hazard rate functions, for each t, extract the 2.5 and 97.5 percentiles of the bootstrapped hazard rate functions to create a confidence band.

In step 3, 1000 replications would provide sufficient accuracy (if the computational or storage costs are a problem).

As an alternative to the bootstrap, the authors might also explore whether the software allows them to extract the design matrix X for the smoothing spline and also the complete variance-covariance matrix, V of the spline coefficients, so that the authors could calculate point-wise SD values analytically as the sqrt of the diagonal elements of X'V'X.

If one of these alternative approaches is taken, then these values should also be used to update the delta method calculations discussed on pages 9-10. That approach which is added to the paper in this revision is correct, the only concern being that valid estimates of the variance of the hazard estimates from the US data are plugged in here.

**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.

**Declaration of competing interests:**

I declare that I have no competing interests.