Reviewer's report

Title: Estimating the BMI-Mortality Relationship: A New Approach

Version: 1 Date: 1 July 2011

Reviewer: Patrick Royston

Reviewer's report:

Comments for authors of Estimating the BMI-Mortality Relationship: A New Approach
Edwin S Wong et al.

The authors describe a new approach, based on fractional polynomials for continuous predictors, to estimating the relationship between all-cause mortality and BMI in samples from the US NHIS annual survey. A major and very welcome contribution is that BMI is retained as a continuous variable in all primary analyses by applying fractional polynomial methodology to it (and to the very important continuous confounder, age). I have the following comments.

Major compulsory revisions

1. It is absolutely critical in a study of the present type that the functional form of the continuous variable of interest (here, BMI) be (a) as close to “correct” as is reasonably possible, given the inevitable limitations of the data, and (b) scientifically plausible. Condition (b) is satisfied in the sense that the authors results appear to agree qualitatively with most other studies in the literature. I believe that more could be done with condition (a). Although FP2 functions are certainly more flexible than quadratic polynomials, they still have their limitations. One helpful approach could be to divide BMI into many (say, 25-40) narrow categories and estimate mortality in each category. This can still be done, with adjustment for the important confounders, in a logistic regression model setting. The results can then be plotted against BMI (i.e. against mean BMI in each category) and compared with the results from the FP analysis. If desired, lowess or other suitable smoothing can be applied to the (inevitably noisy) categorical estimates, but this is not essential. The large sample size in the present study makes such a sensitivity analysis feasible. An alternative to the “small-categories” approach would be modelling BMI by using regression splines (or other flavours of splines), but this is not necessarily a straightforward task unless researchers already have considerable experience of the methodology.

2. The authors nicely demonstrate the existence of a nadir in mortality according to BMI. The nadirs vary according to sex (and in females, also by age). However, they give no model-based estimates of the uncertainty in those nadirs. Confidence intervals for the nadir can quite easily be estimated by the bootstrap method or by the so-called delta method (based on Taylor series expansion). The bootstrap is probably the easier of the two, although it should be possible to use Stata’s remarkable predictnl command to get delta-method based estimates
of the SE of the nadir without (much) mathematical or programming pain.

3. The MFP method as described by Sauerbrei & Royston (1999) [reference 31] incorporates a so-called closed test procedure for selecting an FP model. The “hierarchical” selection method described by the authors (middle of page 7) differs from the closed test procedure, and is known to suffer from an increased type 1 error rate (i.e. can give too many “significant” results). The authors state (page 8) that they used Stata’s mfp command to select an FP model. The default and preferred algorithm for FP model selection in mfp is the closed test procedure. Please clarify which procedure was used, and if not the closed test, please justify.

Minor essential revisions

1. The authors state that they excluded individuals below age 18 years and below BMI 18.5. Please say why.

2. Can the authors please be more explicit as to how they applied “sample adult weights from the NHIS” (page 6) to correct their results for sampling bias, e.g. how was this done using logistic regression with Stata?

3. The authors state that they used 5-year mortality as the main outcome variable. However, it isn’t clear if any times to death were censored in the follow-up interval 0 to 5 years. If so, the mortality results could be biased downwards, since some deaths might have been missed.

4. The authors state (page 7) that “the MFP method also scaled and centered variables in model selection process to improve model fit”. This is not correct. The scaling is to improve numerical stability and the centering is to provide the model intercept with a sensible interpretation.

5. Are the results in these graphs shown in Figure 4 adjusted for confounders? If so, I wonder how the results in the graphs were calculated. For example, the categorical model and the FP2 model for males (bottom left panel) seem to disagree somewhat as to the general level of mortality. The categorical model suggests slightly higher mortality. One has to be careful to ensure comparability of estimates with this sort of analysis.

6. Page 4: “obesity … society faces”: presumably US or Western society, not society globally?

7. Page 4: “Some studies have concluded no relation” – please provide citation(s).

8. Page 5: “endogenously select the best fitting model”: please explain this phrase.

9. Page 5: the NHIS is conducted annually. Why did the authors select data (only) from 1997-2000 for their analysis?

10. Page 5: “BMI missing or … a BMI of over 99.99”: does this throw doubt on the quality of the BMI data?

11. There are several typos in the manuscript that need to be corrected.

12. Figure 1: the relationship between mortality and age looks linear, but the
authors selected an FP1 model in the main analysis. Can they explain this apparent discrepancy?

**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 have no competing interests.