Referee Report: Modelling the mortality for China's oldest-old

Summary

The authors use hierarchical survival-tree methods to fit various parametric models of mortality to Chinse individuals aged 80+ using data from the Chinese Longitudinal Healthy Longevity Survey (CLHLS). The results confirm the importance of covariates (most especially the number of Activities of Daily Living (ADL's) the individuals have difficulty with, gender, and interviewer-rated health status) in assessing mortality, and appears to confirm the age-related slow-down in mortality rates that has been reported in other studies.

Major strengths

The work appears to be rigorously implemented. The question is interesting. Data sample useful & I am not aware that it has been analysed in this way before.

The results appear to confirm an age-related slowdown in mortality rates, documented in many other papers.

Major weaknesses

As far as I am concerned, there appear to be three major weaknesses of the paper:

The first is the writing. Besides a large number of grammatical errors, the paper is too technical in the main body of the paper, and the technical pieces are not well explained. Readers unacquainted with decision-tree modelling (likely to be most readers, and including this reviewer) will struggle. I suggest that the authors (1) explain the broad principles of decision-tree modelling in the main text in a way that is accessible to general readers, and (2) relegate all technical details to appendices. Where formulae are used, the text should provide the insight that the formula uses or captures. As just one example, equation 3.13 defines the Brier Score used to rank alternative models. But the text answers none of the important questions: what is the Brier score, and what is it trying to measure? Why is this a sensible metric to use, and why was it chosen over suitable alternatives? Why should the Brier Score be weighted using the inverse probability of censoring, and what problem is this trying to solve?

Grammatical errors must be corrected. If necessary, the authors should send the paper to a native English-speaking professional copy-editor to ensure that the writing is grammatically correct.

The second issue is the potential role played by sample attrition in driving the results. Table 2 shows the proportion of the sample lost to attrition in each wave. Attrition at this level is without doubt high enough to be driving the entire slow-down observed in the results. I find it extremely puzzling that people older than 80 cannot be found in such large numbers – where are they going and why? Surely the number of people of this age who move house is extremely small. If even a small fraction of attrition is due to unobserved or unrecorded mortality, this could be driving the results (including the apparent slowdown of mortality by age that the authors document). I would therefore suggest that the authors spend much more effort to show that attrition cannot be driving their results. This would include exploring in more detail why attrition happens, and quantifying the likely effects of attrition on the results.

The third issue with the paper is the strangeness of some of the decision trees that result. For instance, Figure 14 shows a survival tree. Firstly, the tree seems to suggest that gender is irrelevant for with ADL

status 3, and ADL status 1 or 2 and interviewer-related health 1. This is counter-intuitive, at best, and suggests that the tree may be ad-hoc data-mining rather than fitting true mortality patterns. Second, ADL status appears at two levels of the tree. While this may capture interactions between the variables, it raises further questions about the extent to which the tree is capturing underlying mortality patterns and not just data mining. Can the set of allowable trees be constrained in some way to demographically-meaningful trees?

Other issues

Introduction – low fertility is by far the primary cause of population aging, rather than 'compounding' it, as stated in the introduction. The intuition is that low fertility causes individuals to be missing across their whole lives, whereas longevity improvement at older ages increases lifespan after, say, age 80 by 2 or 3 years at most. The introduction should make this clear.

Figure 1 should be drawn on a log scale.

Top of page 4 – the literature review should include Zuo at al (2018, PLoS One), which documents mortality slowdown using data from many countries.

Page 4 second paragraph – it may be useful to differentiate between the biological rate of cohort ageing and the demographic rate of cohort ageing (the second includes the selection effect caused by the differential mortality of the frail). The key reference in using differential frailty within cohorts to explain the slowdown is Vaupel et al (1979).

Page 6 – reader is probably not familiar with survival tree modelling, and this description doesn't help. See the major point above.

Page 7 – is the definition of Ex(c) appropriate for high levels of mortality, where the distribution of age at death is likely to be biased toward the earlier part of the year?

Table 1 – May I suggest reparameterizing these equations to preserve a consistent definition of alpha beta lambda and gamma as explained in equation (2.3). Otherwise use different parameter names as these are not comparable in the table.

Page 10 – bottom – what on earth is 10-folds cross-validation? See major point 1 above.

Page 11 – bottom – the null hypothesis 'suggests' nothing. It is a modelling choice, no more and no less.

Page 12 – third paragraph. Why exactly is it not optimal and accurate to calculate the observed and expected number of deaths directly? See major point 1 above.

Page 14 – last paragraph – see major point 1 above.

Page 15 – middle paragraph – ditto. This is where I felt that there were too many technical details in the main text and that the paper would improve if they were relegated to appendices, with the underlying intuitions provided in the main text.

Page 18 – eqn 3.13 see major point 1 above.

Page 19 – this is where I started to be concerned about sample attrition and the effect on observed mortality rates in the sample.

Figure 12 – force of mortality should be on a log scale

Figure 15,16,17 – ditto

Page 39, section 5 – discussion of attrition bias is important but unsatisfying, given the extent of attrition recorded in the paper. I suggest that this section be significantly strengthened to persuade the reader that results are capturing underlying mortality patterns rather than simply sample attrition.