Thesis corrections for “Spatiotemporal modelling of all-cause and cause-specific mortality in England”

Theo Alexander Omar Rashid
2023-01-12

Dear examiners,

Thank you for taking the time to read my thesis “Spatiotemporal modelling of all-cause and cause-specific mortality in England” and for the thoughtful and enjoyable discussion we had during the viva. I attach my response to the corrections.

Kind regards,

Theo Rashid

Required corrections

1. P25 (and P40): While not strictly a paper, a very classical depiction of inequalities in longevity across small areas actually comes from London (see here).

I have added this reference and an explanation of the visualisation to the section on spatial inequality in the UK (PX).

2. P28: please revise the textual distinction made between areal model using spatial neighbourhood structure and those which use nested hierarchy as they are both a specific case of hierarchical models

I have removed the distinction and explained these are both hierarchical models for areal data (PX).

3. P31: Please discuss here or elsewhere that in some cases it is not appropriate to have statistical smoothing and how can someone evaluate if that is the case in a specific situation; we discussed using posterior predictive checks and abplots on death rates

I have added a section explaining that comparisons of fitted and raw data can be used to identify outlier events that we do not want to smooth over (PX).

4. P32: Here or elsewhere please also discuss potential disadvantages in life expectancy as an outcome measure especially as relevant to the assumptions that you need to make in terms of its calculation

I have added detail on the main disadvantage of the interpretation of life expectancy resulting from the assumptions that life tables make about mortality conditions (PX).

5. P48: the idea of exploring population turnover (and therefore a measure of changing composition) is great. The only thing I'd mention is that turnover itself may be an exposure of ill health (some discussion here)

I have now mentioned the "healthy migrant effect", which the suggested paper discusses, in the discussion on chapter 5 (PXX).

6. P54: An important topic not considered in detail is model adequacy. Please add a table summarizing outcomes of model adequacy and consistency checks performed.

I have done so (PX).

7. P56: Please elaborate on your prior choices and robustness of your analyses.
8. P57: Please reconsider your statement that "INLA scales badly" in light of the discussion during the viva and our elaboration that the number of hyperparameters are probably less than 20.
9. P57: It would be useful to see a deeper discussion on the age-specific random walk priors, in light of the J-shape of all-cause mortality rates, and different age profiles of cause-specific mortality rates, as well as a discussion on independence across gender. For example, another way of modelling age (as compared to random walks) is through the use of linear splines (e.g.,

TOPALS): https://www.scielo.br/j/rbepop/a/9szg7XYCXck9dJrKmgBSdrf/?lang=en and the pros and cons would merit discussion.

10. Figure 5.1: what years are used for the country life expectancies? 2002 or 2019?

As stated in the figure caption, the life expectancy estimates “for other countries [are] from World Bank estimates in 2019”.

11. P62: missing from this paper is an evaluation of the uncertainty around life expectancy. This could be measured with a sort of coefficient of variation (SD of all posterior LEs divided over the median LE). The CDC in the US has used 25% as a threshold for poor certainty in some of their papers, though this threshold may be too high.

I opted for a visual evaluation of uncertainty by plotting the credible interval of life expectancy against the median estimates (new figure 5.X). Both of these quantities are in Figure 5.1, but the new figure shows the uncertainty increasing with life expectancy more clearly, which I note on PX. The largest 95% credible interval is 10.6 years, which suggests the SD to median life expectancy ratios (that the CDC use) would all be below 25%. I did not include the CDC’s metric as it is somewhat arbitrary and introducing a new metric sacrificed readability.

12. P73: On the discussion about factors driving lower life expectancy in the US (linked to deaths of despair), a recent paper blames the lack of social safety net systems which would fit well with the authors narrative on the erosion of the welfare system in the UK from 2012 onwards

In this discussion, the deteriorating trends US life expectancy refer to an earlier trend that was studied in Ezzati (2008), who found the trends were driven by increases in mortality from lung cancer and COPD. These are separate to the deaths of despair story, which is discussed in chapter 9, and so I have kept the discussions separate so as not to confuse the stories.

13. P75: please clarify the discussion on spatial analyses of mortality rates at post code level with a greater appreciation of the pitfalls in such.

I have added a footnote on the disadvantages of modelling at the postcode level (PX).

14. P76: Please specify the mean and variance of the BetaBinomial as expressed in your parameters so that the precise choice of parameterisation is clear.

I have done so (PX).

15. P77: there appear to be a few life expectancies in the high 90s or even 100s (City of London), which seems unfeasible. There must be wide uncertainty around these, right? Alternatively, the author mentions a mismatch of population estimates due to changes in adjacent areas.

As with #11, I created a new figure by plotting the credible interval of life expectancy against the median estimates (new figure 6.X). Indeed, there is wide uncertainty around the highest life expectancy. There are also some credible intervals in excess of 30 years for women, which questions the utility and plausibility of these estimates and any future estimates done at smaller spatial units. I have added this to the limitations in the discussion (PX) and the final discussion chapter (PX). As explained in the discussion of chapter 6, some of the highest life expectancies are likely due to a mismatch of population estimates, but with MSOA level estimates in London is excess of 95 years, it is plausible that LSOA level estimates can exceed 100 years.

16. P81: please clarify how the uncertainty is visualised (for instance in Fig 6.5). Is this the distribution at district level of the average posterior mean at LSOA level?

I have added to the figure caption that the distribution is from a density plot of the median life expectancies (Figure 6.X).

17. P82: One thing that is not discussed is the potential for stronger spatial effects in a city like London using LSOAs as compared to the whole of England using MSOAs. The other chapter finds similar results, but segregation patterns over larger areas may differ from smaller areas.
18. P88: please add a description on how cause-of-death is assigned in the UK to clarify the nature of the underlying data; similarly has any uncertainty been considered in the estimated of the contributions of death from each cause and would this be important to consider possibly in future work?

I have added a further details of how cause-of-death is assigned and how the ONS use computer-based selection algorithms to improve consistency (PX).

The quality of cause-of-death assignment was also mentioned by the reviewers of the paper that resulted from chapter 8, and so we expanded on our limitation section for that paper. The UK has the highest rating based on completeness and a low share of deaths assigned to implausible and ill-defined codes, and validation study has found good agreement between cause of death assignment and an expert panel for prostate cancer diagnosis. I have added these details, which are largely taken from the reviewed paper, to the section of the discussion describing issues with cause of death assignment (PX).

19. P88: It is not clear what the “ICD-10 code in the first position of the death record” is for the classification of deaths for neonates without underlying cause of death. Does this refer to contributory causes?

I have added further details on the death record data and cause-of-death assignment to clarify this statement (PX).

20. P89: It is not entirely clear on where ill-defined deaths go (R chapter mostly). The chapter does not mention anything, and Table D.1 mentions those deaths being included in the residual causes (but are not listed in those). There’s also a mention to a redistribution without any details.

I have now clarified in the main text how the ill-defined deaths were assigned to the residual groups (PX).

21. P91: the descriptions on Arriaga’s method are very hard to follow without equations, and please lay out key assumptions also in the main body of the thesis.

I have supplemented the text with the equations and reiterated some details from Arriaga’s method calculation in the appendix to the main text (PX).

22. P91: if the causes of death cover all causes (I don't see anything missing), why is the rescaling of cause specific mortality needed? (there's a mention in P103 to the "total mortality rule" which I assume refers to this, but it is unclear)

The rescaling of mortality was specifically for the cause-specific contribution to mortality analysis. I have added detail to the text explaining this and that it required so that the contributions sum exactly to the life expectancy differences (PX). This is reiterated in the footnote on PX.

23. P97: what may be reasons of the far more common contribution of lung cancer to inequality among women compared to men? (this is somewhat talked about in chapter 8)

I have added a sentence on inequality to the section of the discussion in chapter 7 on the delay in the peak of smoking for women (PX).

24. P109: please better justifying the need for this project. There is not much in the introduction about this.

Chapter 8 has turned into a paper, which actually had an even shorter introduction than in this thesis. Nonetheless, I have provided some further context from the paper explaining the lack of cancer mortality data available for small areas in England (PX).

25. P111: as with our earlier comment above on ill-defined causes, it isn't clear here how they are treated for this analysis.

I have added a sentence explaining that ill-defined diseases were reassigned as in the previous chapter (PX).

26. P112: The author indicates there are very few studies looking at geographic patterns and cancer. Spain has been producing reports on that for some time (See here, the newest one is in English, direct link to the newest one here)

I have now referenced this report on Spain and Portugal as well as a similar atlas from Australia (PX).

27. P117: reword the definition of posterior probability of a true decline

I have reworded the sentences to avoid the phrase “true decline” (PXX, LXX).

28. P125: For the discussion on epidemiologic transitions, I suggest this and this for some nuanced discussions. Injuries have definitely been a weak spot of the theory.

Thank you for providing these references. I have used these to provide further detail on the limitations of the epidemiologic transition theory (PX).

29. P145: please provide a full justification of equation A.1, and under what assumptions death probabilities conditional upon survival to age x can be derived from cross-sectional, empiric mortality rates.

I have extended the derivation to show how equation A.1 can be derived intuitively, and have added that the results assume the mid-year population is used to approximate the number of person-years lived (PX).

30. P145: please provide a worked example to allow readers to build intuition into the approach, we suggest when n=1.

Equation A.1 is not intuitive, and there is no text in Preston (2001) (where all the life table detail is taken) or other sources that tries to build intuition - it is simply a result of algebraic manipulation. By providing further detail on the intuitive derivation of equation A.1 in #29, the reader is now better-placed to understand how the result has arisen.

31. P146: please explain in words how the equations ensure that the life table is closed in the sense that everyone must die at some point

I have extended the derivation to show how the expression for the number of survivors in the open-ended age interval has been derived intuitively using the fact that the number of deaths in the open-ended interval is equal to the number of people who survived to the final age group (PX).

32. P146: please provide reasonable detail and assumptions behind the Kannisto-Thatcher approach so the thesis is more self-contained.

I have added more detail on the Kannisto-Thatcher method, including the assumption that the method assumes the probability of dying is a logistic function of age (PX).

33. P146 Section A.2: please provide a more detailed justification with greater attention to the underlying assumptions that lead to equation A.6. We did not follow details such as “subtract the probability of surviving”. Further, it seems the construction uses D^i_x from some year t to construct a hypothetical cohort that survives to cause i alone. However there are competing deaths and if these had not occurred, D^i_x would be larger for old ages x. It seems the underpinning are strong and these render interpretation of the calculated cause-specific life expectancies challenging.

I have clarified the phrase “subtract the probability of surviving” in section A.2 (PX). I have added further detail on what the terms in Arriaga’s method represent (PX).

The construction of the age-specific ${}_n\Delta_x$ relies only on differences in life table quantities from all-cause mortality, i.e. those used to calculate the life expectancy for each population, rather than any hypothetical cohort exposed to cause i alone, which would be used to construct cause-specific life expectancies. For the cause-specific contributions, the main assumption is that ${}_n\Delta_x^i$ are proportional to the differences between the populations in cause-specific death rates in that age group. These are not cause-specific life expectancies, but cause-specific contributions to differences life expectancy (calculated from all-cause death rates) between two populations. I have reworded the text to clarify that these are the same life table quantities from the all-cause death rates used earlier in Appendix A (PX).

Other

1. Add Marta’s paper.
2. Cite latest paper for chapter 8.

Chapter 8 turned into a paper, which was released on the same day as the viva. For completeness, I have edited the first sentence of the chapter to reference the paper (PX).