Questions and answers on "Applying survival models to pensioner mortality"

During the discussion on 25th February 2008 a number of points were made in response to the paper. Some of these points were wrong, while others were potentially misleading. The author did not have the opportunity to reply fully to them all on the night, and this note seeks to do so here.

1. "Independence [of observations] is not important in practice"

The assumed independence of the events being modelled is a critical foundation for even the most basic of statistical models. To ignore the independence assumption is therefore negligent. This is particularly so for annuity portfolios, since it is very common for people to have multiple policies. There are two means of dealing with this: deduplication (the preferred option) or fitting a model with an over-dispersion parameter. Since we are building a model of mortality at an individual level, we can use deduplication to ensure broad independence of observations. Where a model only has aggregated data, as is the case with historical CMI data series, an overdispersion parameter must be used.

2. "Non-linearity is only important for centenarians"

This comment is perhaps intended to apply to the progression of mortality with age. Figure 4 (reproduced below) shows that this is not true: non-linearity is quite obvious after age 90. Furthermore, the failure of a model to follow the non-linear pattern affects not just the older ages, but also distorts the fit at younger ages as well. The left-hand panel of Figure 4 also shows that the linear assumption results in systematic under-fitting from age 70 onwards, so non-linearity is clearly not just important for centenarians.

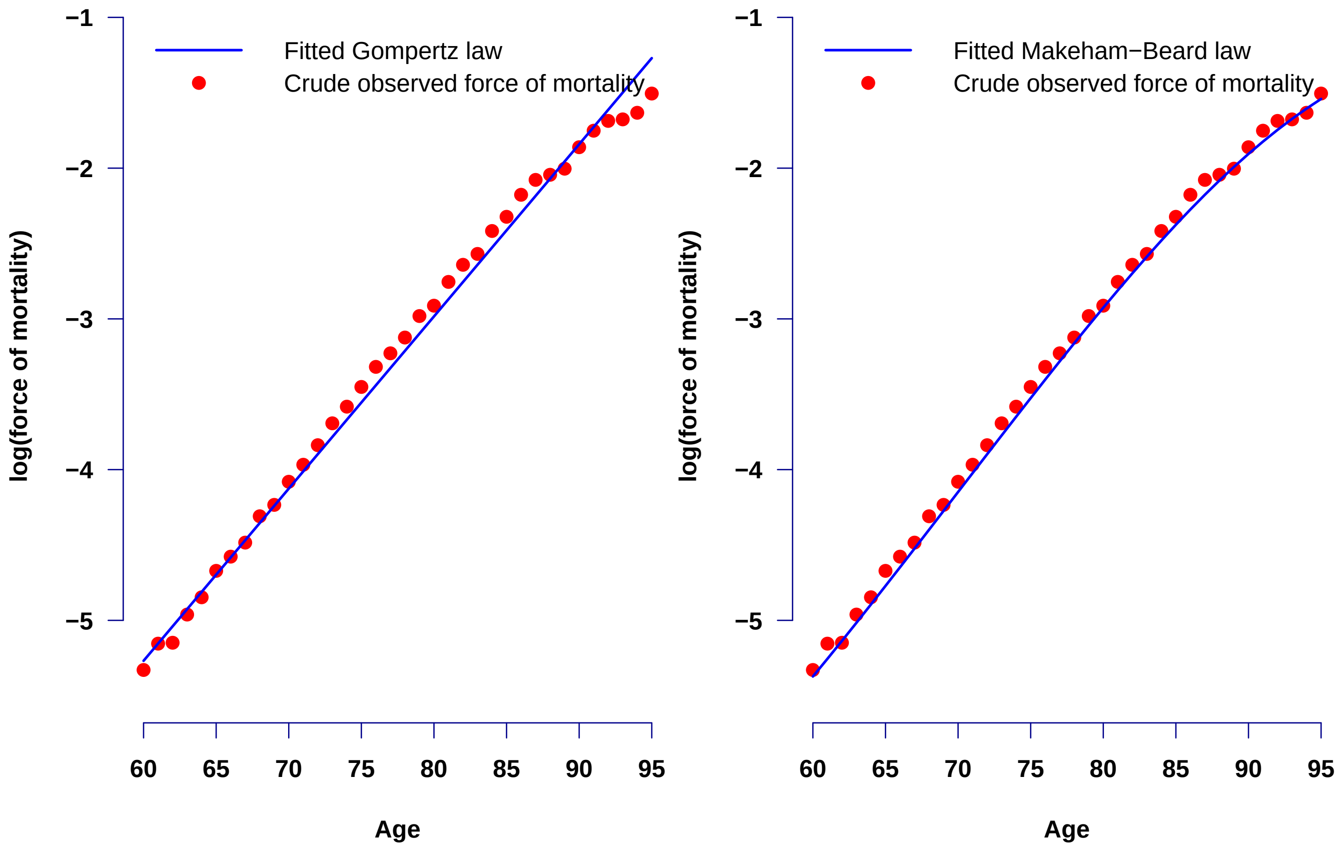
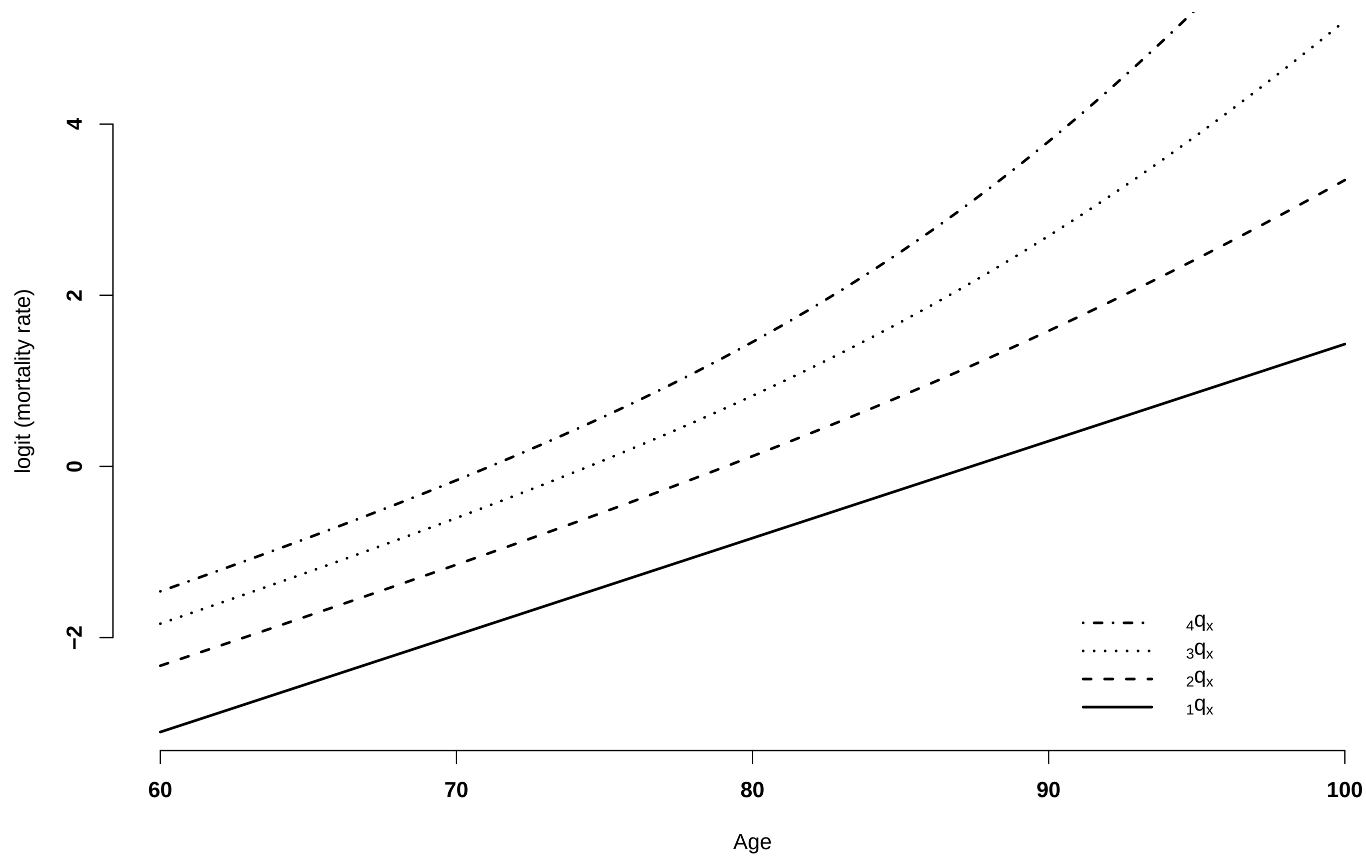


Figure 4. Gompertz model (left) and Makeham-Beard model (right).

3. "Generalised Linear Models (GLMs) are better than survival models"
This claim was not backed up, nor was it stated which type of GLM was meant. For example, the comment might have referred to contingency tables, or Poisson models for the number of deaths in a group. In this case information on individual lives is lost during the grouping, and the need to maintain a minimum expected number of events in a group severely constrains the number of risk factors which can be modelled. Alternatively, what might have been meant was logistic regression, which parallels survival models in that it is a model of individual mortality. q_x is usually logit-linear above age 60 and logistic regression works well, as reported by Richards and Jones (2004). However, if q_x is logit-linear then ${}_nq_x$ is not where $n > 1$, so logistic regression can only be used like this for a single year's data. The figure below shows how, if q_x is suitable for logistic regression, then it is unsuitable for a multi-year model of mortality.



4. "GLMs are less parametric than survival models"
This is incorrect. The logistic regression used by Richards and Jones (2004) is a GLM, and it contains parameters which are the direct analogues of the parameters in Table 13 of the paper. When fitting an equivalent mortality model, the GLM parameters in logistic regression have the same sign and size as those in the survival model. The main difference is that survival models are more powerful, use all of the data and tend to produce smaller standard errors around the parameter estimates.

5. "Splines can be used in GLMs"
True, but there is nothing to stop them being used in survival models. One advantage of splines lies in handling non-linear relationships (which is important for more than "just centenarians" as described earlier).

6. "GLMs can be used model mortality monthly"
True, but why stop at monthly? A survival model deals with mortality *daily*, and can therefore use every single piece of data you have.

7. "The automated searching algorithms over-segment the population"
This is incorrect. Survival models cannot "over-segment" without this becoming patently obvious in either the AIC or the standard errors on the parameter estimates. Indeed, the automated procedures outlined in the paper do not "exacerbate flaws"— on the contrary, the use of the AIC as a target ensures that over-segmentation cannot happen.

8. "Smoothness is unimportant"
This is a most unusual claim, as most actuaries would believe this is the whole point of graduation. It is probably not putting it too strongly to say that smoothing is what is behind nearly all of statistics, from calculating a mean to fitting a complex 2-D surface to a mortality array. Comments were also made about so-called "self-prophesying models", which seem to be a warning about fitting parameters to small sub-groups then re-aggregating. This is to misunderstand the methodology used in this paper, which maps lives to lifestyle via postcode *before* any model is fitted, not afterwards. In any case, the smallest lifestyle group in the model contains 416,937 life-years of exposure and 13,545 deaths, which we think most people would agree is quite large enough.

9. "Generalised Additive Models (GAMs) are the way forward"
GAMs are indeed highly flexible, and can provide a good fit in the presence of nonlinear relationships. However, this flexibility must be used with great caution. Look at the right-hand panel of Figure 4 — why use numerous B-splines when a simple, four-parameter mortality law will fit the data just as well? It is important to evaluate whether the added complexity of generalised additive models is really required, and this is usually not the case for pensioner mortality. Given a comparable fit of the models, the simpler mortality-law approach is preferable to the more complex generalized additive model. For example, a GAM with B-splines centred at five-year age intervals would involve substituting eight spline parameters for each of the six used for age-related mortality in Table 13 in the paper. This would result in a net increase of 42 parameters ($42 = 8 \times 6 - 6$), and we strongly doubt the model would be improved by enough to warrant such a large increase in complexity, even allowing for any reduced effective dimension due to smoothing.

10. "The proposed structures are new and unproven"
Proportional hazard models have been around since they were proposed by Cox in 1972, i.e. they are 36 years old at the time of writing. Cox's paper is one of the most cited in mathematical literature, with around 600 citations a year. Furthermore, modelling the force of mortality lies behind all recent CMI table graduations. As for the mortality laws themselves, the youngest mortality law used in the paper dates back to 1932, while the oldest has been around since 1825.

11. "Mortality-based postcode groupings are better than Mosaic"
No evidence was advanced for this assertion. We have fitted models based on postcode grouping, models based on Mosaic, and models with both simultaneously. In every case we have found Mosaic (or Acorn) to be a much more powerful predictor of mortality than any other postcode-based grouping. Models using Mosaic or Acorn type are also easier to deal with than the hundreds of thousands of post codes, in addition to passing the significance tests.

12. "Postcode is unnecessary, pension size is all you need"
This rather contradicts the point above, but, to see that this is not so, consider a pension of 3,800 p.a. to a male aged 65. This puts the pensioner in the highest of the three pension bands in Table 11. However, the difference between this pensioner being in a higher or lower lifestyle class is 7.1% for an escalating pension. Since a typical pricing margin for an annuity or bulk buy-out is around 5%, being able to use postcode is essential for profitable pricing. To put it another way, a company writing annuity business ignoring postcodes will lose money if it is competing against companies which do.

Table 11. Impact of pension size and lifestyle

Gender	**Pension size**	**Lifestyle**	e_{65}	$\bar{a}_{65}^{5\%}$	$\bar{a}_{65}^{2.5\%}$	**Change in** $\bar{a}_{65}^{5\%}$	**Change in** $\bar{a}_{65}^{2.5\%}$
Female	Highest	Upper	22.88	13.26	17.05	n/a	n/a
Male	Highest	Upper	20.23	12.23	15.43	-7.8%	-9.5%
Male	Highest	Lower	18.56	11.50	14.34	-6.0%	-7.1%
Male	Middle	Lower	17.06	10.83	13.36	-5.8%	-6.8%
Male	Lowest	Lower	15.62	10.12	12.37	-6.6%	-7.4%
Overall						-23.7%	-27.4%

Stephen Richards Consulting Ltd,
28th February 2008.