Predictive Mortality Graduation and the Value At Risk: A Bayesian Approach

M. Mendoza
Comisión Nacional de Seguros y Fianzas and
Departamento de Estadística, ITAM

A. M. Madrigal
Comisión Nacional de Seguros y Fianzas

E. Gutiérrez-Peña
Departamento de Probabilidad y Estadística, IIMAS-UNAM

January 7, 2000

Abstract

Graduation and overestimation of death rates play a key role in the construction of mortality tables. The usual actuarial methods deal with each of these two aspects separately. Moreover, the statistical nature of these problems is typically ignored. In this paper, a method to produce mortality tables is proposed which simultaneously takes both issues into account. The approach is entirely statistical and allows the user to select a table providing a precisely defined level of protection against deviations in mortality. The problem is analyzed from a predictive perspective within the Bayesian framework and the solution makes use of the concept of value at risk.

Some key words: Actuarial methods; Bayesian inference; Mortality models; Linear regression; Predictive distribution; Overestimation; Value at Risk

1 Introduction

Mortality tables play a key role in life insurance. Past experience is used to estimate the probability of death for every person who buys a life policy. There are a number of factors which are supposed to affect this probability, age being the most typically considered. Other characteristics include sex, clinical records, occupation, and life habits such as smoking or the practice of sport activities. In any case, the relevant probabilities are estimated from demographical records of the target population – or other similar data– and on this basis premiums and reserves are calculated. A typical mortality table is an arrangement of death probabilities q_x, according to age X, of individuals. Usually, true probabilities follow a positive monotonic trend (older individuals are more likely to die than younger ones). On the other hand, estimated probabilities show a similar pattern, but affected by sampling errors. The problem then is to produce an estimated table showing the theoretical monotonic trend. Thus, a smoothing technique is required to obtain this graduated table (see London 1985, for example).

Another relevant issue arising in actuarial practice, due to the conservative nature of insurance systems, is the need for overestimated death probabilities. The basic idea is to increase the graduated probabilities in order to reduce the risk of an insufficient reserve. Throughout this paper, we shall refer to the arrangement of such probabilities as a *modified* table.

For many years, graduation and overestimation have been implemented ignoring, for the most part, the statistical nature of this kind of data, although some contributions have recently appeared with a much more statistical content (Forfar *et al.* 1988, Renshaw *et al.* 1997 and Hand and Renshaw 1998 among others). These authors, however, are only concerned with the graduation problem.

Concerning overestimation, it is worth noticing the predictive nature of the problem. The purpose is to develop robust insurance systems in the sense of preventing insufficiencies caused by extreme observed mortality rates. Thus, one should focus on the future behavior of the observed mortality rates in order to obtain modified tables whose probabilities of death will be exceeded by the future observed rates only with a reasonably low probability.

In this paper, a Bayesian approach to modeling is adopted to deal with both, graduation and overestimation issues, in an integrated manner. The main idea we wish to put forward here is that both problems should be solved by means of a single principle involving the joint predictive distribution of the observed mortality rates.

In Section 2, a real set of data provided by Mexican insurance industry is introduced and a preliminary statistical analysis is carried out. These data will be used throughout this paper to illustrate our proposal. Section 3 is devoted to the modeling stage where the main outcome is a joint predictive distribution for the observed death rates which takes into account the relationship between the probability of death and age. This predictive approach is used to produce a family of modified tables. In addition, a procedure to select a particular modified table is advanced in Section 4 where the predictive distribution for the aggregated number of deaths, corresponding to the specific insured population, is analyzed for this purpose. Finally, in Section 5 some generalizations and related topics are discussed.

2 Mexican Mortality Data

A typical set of data considered when a mortality table is to be produced has the following structure. Let $\mathcal{D} = \{(X_i, E_{x_i}, D_{x_i});\ i = 1, \ldots, N\}$ where X_i is the i-th observed age, whereas E_{x_i} and D_{x_i} represent the number of exposed individuals and deaths corresponding to X_i. For any age X, the observed death rate is defined as $\widehat{q}_x = D_x/E_x$. Usually, the data includes ages ranging from 12 to 99 years. Table 1 exhibits a real data set, provided by the Mexican insurance industry, that will be analyzed throughout this paper.

Figure 1 shows the frequency distribution for the age of the 6,678,352 exposed individuals. Most of the cases lie between 20 and 65 years and their distribution is clearly asymmetric. As for the age at which each of the 23,918 recorded deaths occurred, Figure 2 exhibits the corresponding histogram. The asymmetric pattern is less obvious here, suggesting that the probability of death is not constant across age. In fact, from Figure 3, it is clear that observed death rates follow an increasing trend and show considerably more variability for older ages, where the samples sizes are smaller. The usual actuarial practice is to describe this behavior in terms of a simple curve such as the well-known Gompertz or Gompertz-Makeham model (Benjamin and Pollard 1980, Bowers *et al.* 1986 and London 1985, among others).

From a statistical point of view, however, the different levels of variability across ages must be taken into account when selecting a class of models. To this end, in the remainder of this section we shall examine, for each age, the marginal variability of the corresponding death rate.

For the sake of simplicity, let us assume that, for any age X, and given the exposure frequency E_x, the number of corresponding deaths follows a *Binomial* distribution with probability q_x, the true death rate, and index E_x. This assumption is not uncommon and is derived from a more basic hypothesis according to which all individuals of the same age die independently with the same probability. This structure is not always warranted, particularly in the context of multiple policies, where overdispersion problems may arise. Nevertheless, in this paper the usefulness of any given model will be assessed in terms of its ability to reproduce the behavior of the observed data.

From a Bayesian point of view, we must elicit a prior distribution for q_x and then derive the associated predictive distribution for the number of deaths to be observed, D_x^*. As a consequence, since the future observed death rate is proportional to D_x^*, the predictive distribution for $\widehat{q}_x^*$ is readily available. At this point, if we decide to produce inferences based mainly on the observed data, we can adopt a prior representing a vague state of knowledge about q_x. Here, we shall use the reference prior $Beta(0.5, 0.5)$ as suggested by Bernardo(1979) (see also Bernardo and Smith 1994, for a complete discussion of the concept of reference prior as well as the more general idea of non-informative priors). The resulting posterior predictive distribution for the observed number of deaths D_x^* is $Beta\text{-}Binomial(D_x + 0.5, E_x - D_x + 0.5)$. The corresponding predictive interval, with probability content 0.95, for each future observed death rate is shown in Figure 4. There, the higher uncertainty involved in the estimation of the true death probabilities for older ages becomes apparent. This clearly shows the wide range of possible models one could entertain when analyzing these data. Moreover, there are some extreme observations which cannot be described through a simple, monotonic model. These (rather few) observations might be considered as outliers. We note that these observations basically only increase the variability and do not seem to have a significant effect on the trend.

3 The Model

In order to describe the relation existing between the observed death rates and the corresponding ages we could use again the model

$$\widehat{q}_x \sim \frac{1}{E_x} Binomial(q_x, E_x)$$

with the addition of a function $q_x = u(X)$ where $u(X)$ is non-decreasing on X. Other possibility is to apply a transformation to the observed death rates $Y_x = h(\widehat{q}_x)$ and then to describe the random variable Y_x with a probability model

$$Y_x \sim F(\cdot|\theta_x)$$

where F belongs to a convenient family of distributions and the unknown parameter vector θ_x is supposed to depend on X through an appropriate function $h(X)$.

Both approaches are well known in the statistical literature and the first leads, in a natural way, to the idea of generalized linear models where selection of the function $u(X)$ amounts to the choice of a link function. Some authors have explored these ideas for mortality graduation from a frequentist non-predictive point of view (Renshaw 1991, Haberman and Renshaw 1996 and the references therein). Here, we shall consider the idea of transformation within the Bayesian framework stressing the predictive nature of the problem.

After inspection of Figure 3, the data in Table 1 was transformed using a logistic transformation. The resulting response (*logit*) variable is defined as

$$Y = \ln\left(\frac{\widehat{q}}{1-\widehat{q}}\right).$$

This transformation was applied to all the observations, with the only exception of age 16 since the corresponding observed death rate is zero. Figure 5 shows a clear linear trend for the transformed data. The proposed model is then

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where $\epsilon_1, \ldots, \epsilon_N$ are i.i.d. random variables following a *Normal* distribution with zero mean and variance σ^2. In matrix notation we have

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}).$$

For convenience we shall work in terms of the precision parameter $\tau = 1/\sigma^2$.

Within a Bayesian framework, the joint reference prior for the parameters β and τ leads to the posterior distribution given by

$$P(\beta \mid \tau, \mathcal{D}) = St(\widehat{\beta}, \tau^{-1}(\boldsymbol{X}^t\boldsymbol{X})^{-1}, N)$$

and

$$P(\tau \mid \mathcal{D}) = Gamma(a, b\widehat{\tau}^{-1})$$

where $St(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, k)$ stands for a multivariate *Student's t* distribution with mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and k degrees of freedom. In addition, $\widehat{\beta}$ and $\widehat{\tau}$ are the maximum likelihood estimators of β and τ respectively whereas $a = (n-2)/2$ and $b = n/2$. As a consequence, it can be easily shown that the posterior joint predictive distribution for a vector $\boldsymbol{Y}_*$, corresponding to a matrix $\boldsymbol{X}_*$, is given by

$$P(\boldsymbol{Y}_* \mid \boldsymbol{X}_*, \mathcal{D}) = St(\boldsymbol{X}_*\widehat{\beta}, \widehat{\tau}^{-1}(\boldsymbol{X}_*(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}_*^t + \boldsymbol{I}), N).$$

If we now define $\boldsymbol{x}_*^t = (1, X_*)$ we have that $\widehat{\mu}_* = \boldsymbol{x}_*^t\widehat{\beta}$ describes the relation between the expected value of Y_* and the age X_*. Since the t distribution is symmetric, $\widehat{\mu}_*$ coincides with the posterior mean as well as with the posterior median of Y_*. Applying the inverse (non-linear) transformation

$$h^{-1}(\widehat{\mu}_*) = \frac{\exp(\widehat{\mu}_*)}{1 + \exp(\widehat{\mu}_*)}$$

what we obtain is no longer the posterior mean of $\widehat{q}$ but it still coincides with the posterior median of this random quantity. Since, in general, quantiles are invariant under smooth, monotonic transformations such as $h(\cdot)$, we can easily derive the predictive p-quantile of $\widehat{q}$ for every age by simply obtaining the corresponding quantile of the posterior predictive distribution of Y_* and then applying the inverse transformation. Figure 6 shows the fitted –central or equivalently, median– model for the observed death rates. It also includes a 0.95 predictive band for $\widehat{q}$ whose limits are given by the modified models derived from the 0.025 and 0.975 quantiles.

A first approach to the assessment of a reasonable level of overestimation is readily available from this procedure. A modified model which uses the corresponding p-quantile provides estimations which guarantee that future observed death rates can be expected to produce values above that level, for every single age, only with probability $1 - p$.

Let us recall that the total number or claims depends on both the probabilities of death and the particular age structure of the insured (exposed) population. Thus, a more realistic protection level is obtained upon examination of the posterior predictive distribution for the aggregate number of claims.

A solution which allows selection of the modified model according to this idea in terms of a single global measure of risk will be introduced in the following section. It will be shown that this measure can be interpreted as a *value at risk* (see, for example, Grayling 1997).

4 Prediction of the Aggregated Number of Claims

In real life applications, when using mortality rates for financial calculations, there are two components that must be taken into account: the mortality rates themselves and the demographic profile of the specific insured population.

For a set of death probabilities $\{q_x\}$ and given the set of exposure frequencies $\{E_x\}$, the expected aggregated number of claims is $A = \sum_x q_x E_x$. In this expression, the observed number of deaths (claims) is recovered by using $\{\widehat{q}_x\}$ instead of $\{q_x\}$. However, the quantity of interest is the aggregated number of claims *to be observed* in the future, $A^* = \sum_x \widehat{q}_x^* E_x$. It should be noticed that the predictive distribution for A^* can be derived from that of $\{\widehat{q}_x^*\}$, although it is not available in closed form. Nevertheless, a sample from this distribution can be easily obtained via simulation.

The posterior predictive distribution for A^* can be used to assess the overall protection provided by an specific (modified) model when applied to a particular insured population. Let $\{q_x^+\}$ be the death rates corresponding to the modified model and $A^+ = \sum_x q_x^+ E_x$. Then, $\alpha = \Pr(A^* > A^+|\mathcal{D})$ provides a global measure of the specific risk to be faced when using this model for this particular population. In fact, A^+ can be regarded as the value at risk of level α for this problem and, more importantly, for a given set of exposure frequencies there basically exists a one-to-one correspondence between the value of A^+ and the corresponding modified model. Thus, we can select the appropriate model by choosing a reasonable global value at risk.

In order to illustrate the type of results which can be obtained if we adopt this proposal, we now analyze the Mexican mortality data described in Section 2. Figure 7 exhibits the posterior predictive distribution of A^*, assuming the same exposure distribution of the original data in the example; the graph is based on a simulated sample of size 30,000 from the joint posterior predictive distribution of $\{\widehat{q}_x^*\}$. The mean and standard deviation are 24,012 and 1,569 respectively. In addition, this distribution has a *Normal*-like shape and the 0.025 and 0.975 quantiles are given by 21,069 and 27,240. Furthermore, the observed number of aggregated deaths can be expected to be almost surely not greater than 27,921 (the 0.99 quantile). Now, if we compute A^+ for the modified models corresponding to the levels $p = 0.7, 0.8, 0.9$ and 0.95, we get the figures in Table 2. It can be then seen that a modified model of level 0.7 suffices to obtain a protection of approximately 0.975 probability in the scale of the aggregated deaths (and a *value at risk* of 27,234). The most important consequence of this result is that we do not need to use an extremely conservative model to get a reasonable level of global protection. This fact is not surprising if we notice that a modified model assumes the same extreme probabilistic deviation across ages, whereas A uses the exposure frequencies as weights to produce the total number of deaths; in other words, in this latter case not all deviations are regarded as equally important.

5 Concluding remarks

The method introduced in this paper permits the production of mortality tables by means of an algorithm which simultaneously deals with the graduation and the over-estimation issues. The method recognizes the statistical nature of the problem and makes use of a predictive approach from a Bayesian point of view. The basic steps are (i) the fitting of a regression model to the transformed observed death rates; (ii) the calculation of the joint posterior predictive distribution for the death rates *to be observed* in the future; and (iii) the selection of a modified model through the computation of the associated value at risk with respect to the posterior predictive distribution for the aggregated number of claims to be observed given a specific set of exposure frequencies. It must be stressed that the method is quite general in the sense that, first, other data transformations might be used instead of the logistic; secondly, other monotonic structures (predictors) might be used to describe the relationship between age and the unknown death probabilities. All statistical concepts involved have been widely discussed and are well established in the literature. In addition, it must be pointed out that sampling-based inferences are currently a standard practice, specially in the Bayesian context, where it is often the case that calculations are not feasible in closed form.

It should be noticed that the posterior predictive distribution for the aggregated number of claims in this paper was produced for the same exposure frequencies used to fit the model. However, in practice it would be more convenient and useful to calculate this distribution for other, different, frequencies. This might be case if, for example, the past mortality experience is to be used to explore the consequences of a new risk selection strategy.

Our proposal uses the aggregated number of claims as the key quantity to be considered in order to assess the appropriate level of protection provided by a mortality table. It must be recognized, however, that the loss suffered by an insurance system is not the number but rather the aggregated monetary *amount* of claims. The proposed procedure is equivalent to the one which considers the amount of claims only if the insured sum is assumed to be the same for each policy. A more realistic treatment would involve a statistical description of the variation among insured sums. This extension is now under investigation and some results will be reported elsewhere.

Acknowledgements. This work was partially supported by the Sistema Nacional de Investigadores, Mexico. The third author wishes to acknowledge partial support from PAPIIT grant # IN101198.

References

Bernardo, J.M. (1979). Reference Posteriors Distributions for Bayesian Inference (with discussion). *Journal of the Royal Statistical Society B* **41**, 113-147.

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* Chichester: Wiley.

Forfar, D.O., Mc Cutcheon, J.J. and Wilkie, A.D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries.* 115, 1-135.

Hand, D.O. and Renshaw, A.E. (1998). *Statistics in Finance.* London: Wiley.

London, D. (1985). *Graduation: The revision of estimates.* London: Actex Publications.

Renshaw, A.E., Haberman, S. and Hatzopoulos, P. (1997). On the duality of assumptions underpinning the construction of life tables. *Astin Bulletin.* Vol. 27, no. 1, 5-22.

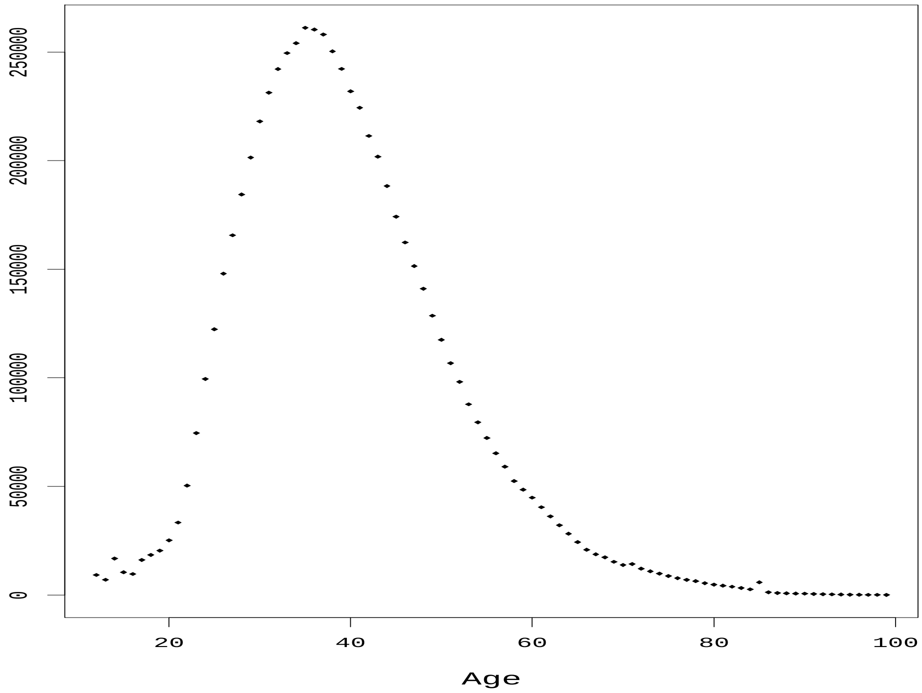

Figure 1: *Exposed individuals*

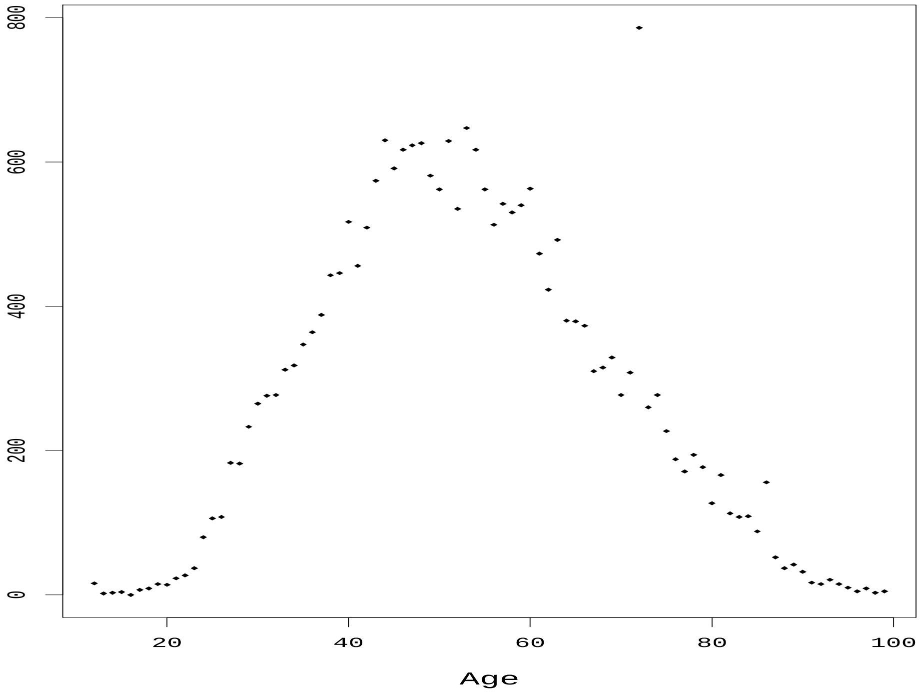

Figure 2: *Observed deaths*

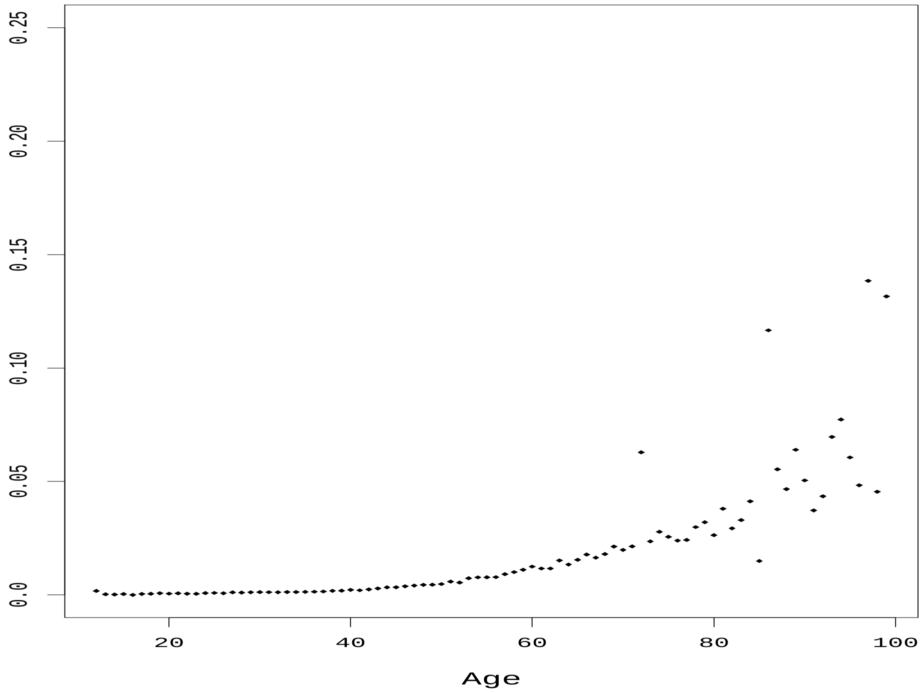

Figure 3: *Observed death rates*

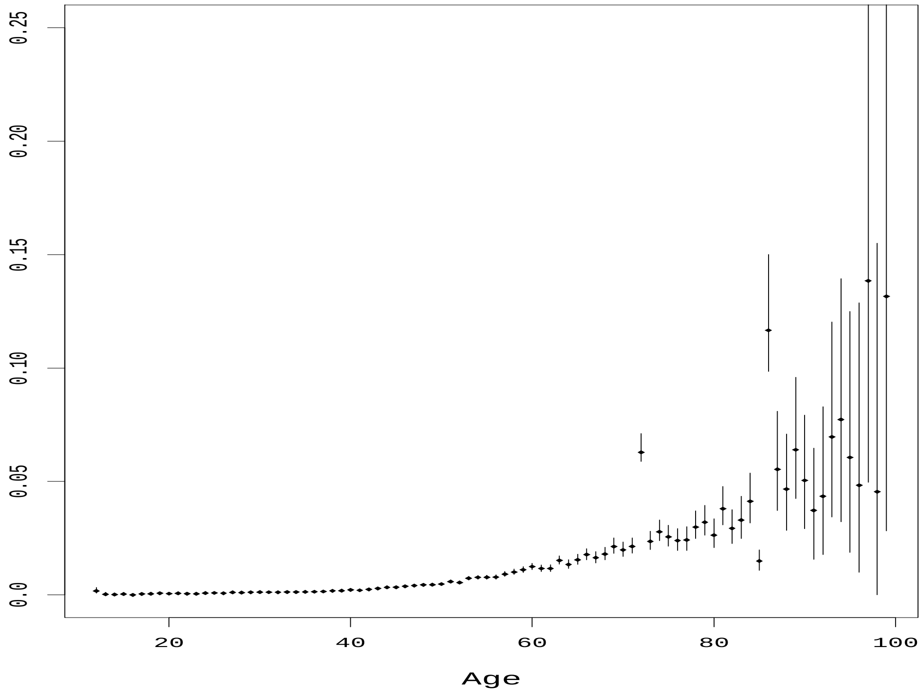

Figure 4: *Beta-Binomial intervals for* q_x

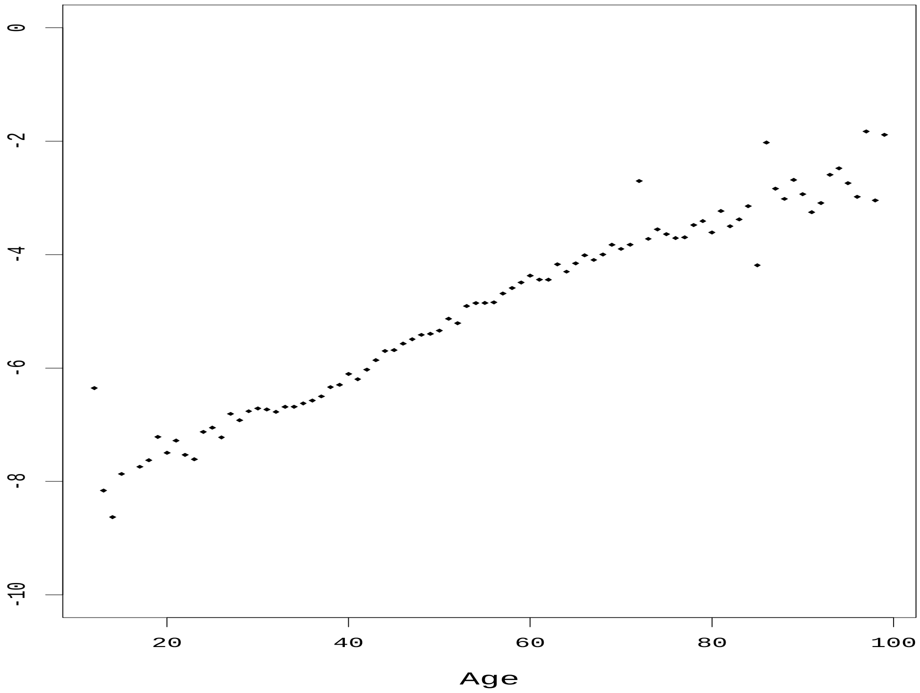

Figure 5: *Transformed death rates*

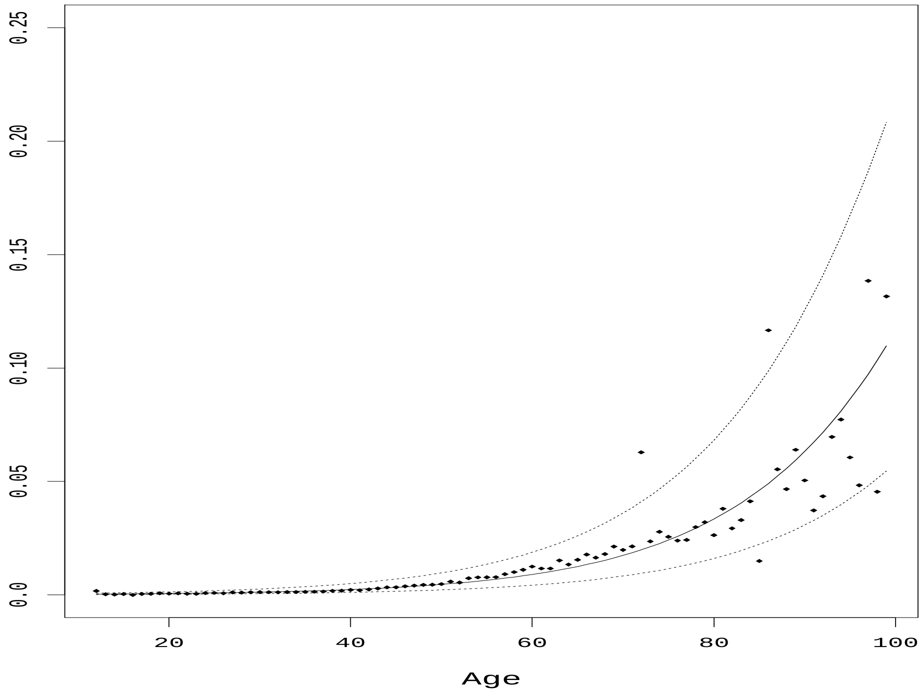

Figure 6: *Fitted model and 95% predictive band*

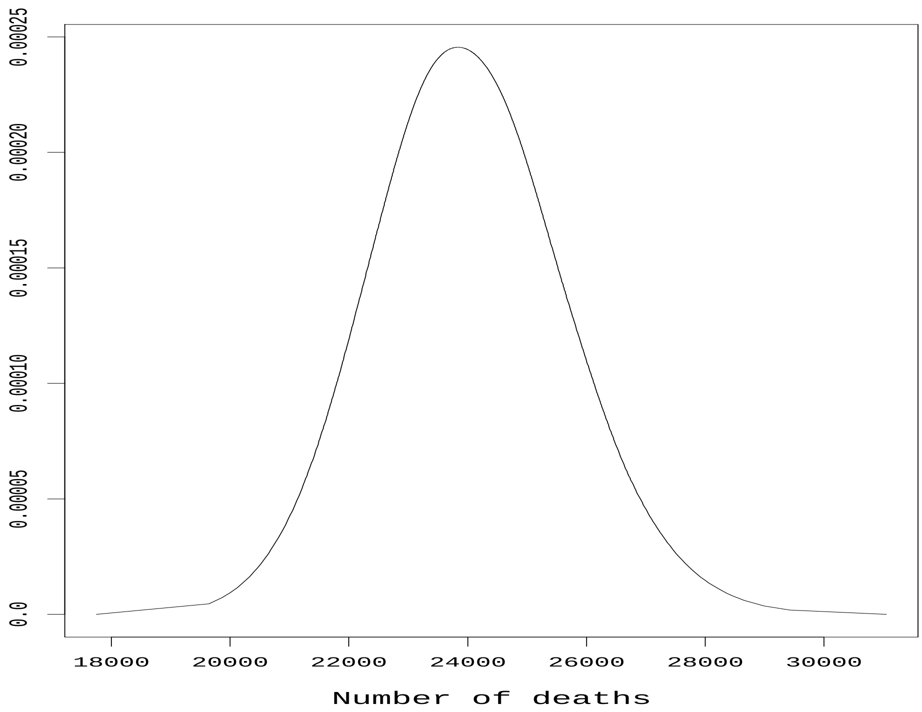

Figure 7: *Predictive distribution of aggregated number of deaths*

Table 1: Exposed individuals and Deaths

Age	Exposed Individuals	Deaths	Age	Exposed Individuals	Deaths	Age	Exposed Individuals	Deaths
12	9203	16	42	211408	509	72	12116	786
13	7001	2	43	201843	574	73	10888	260
14	16768	3	44	188313	630	74	9822	277
15	10445	4	45	174223	591	75	8752	227
16	9654	0	46	162347	617	76	7762	188
17	16102	7	47	151459	623	77	6973	171
18	18442	9	48	141042	626	78	6392	194
19	20408	15	49	128595	581	79	5440	177
20	25173	14	50	117505	562	80	4758	127
21	33359	23	51	106745	629	81	4286	166
22	50374	27	52	98154	535	82	3799	113
23	74561	37	53	87818	647	83	3222	108
24	99480	80	54	79525	617	84	2588	109
25	122342	106	55	72288	562	85	5841	88
26	147954	108	56	65234	513	86	1259	156
27	165651	183	57	59083	542	87	913	52
28	184425	182	58	52445	530	88	775	37
29	201460	233	59	48508	540	89	635	42
30	218085	265	60	44817	563	90	618	32
31	231340	276	61	40417	473	91	448	17
32	242170	277	62	36168	423	92	337	15
33	249538	312	63	32128	492	93	291	21
34	254078	318	64	28208	380	94	186	15
35	261217	347	65	24342	379	95	160	10
36	260378	364	66	20805	373	96	101	5
37	258114	388	67	18746	310	97	60	9
38	250332	443	68	17329	315	98	64	3
39	242261	446	69	15277	329	99	35	5
40	231921	517	70	13819	277			
41	224367	456	71	14268	308			

Table 2: Quantiles and Predicted number of deaths

Modified Model (level)	Predicted number of deaths
0.70	27234
0.80	30620
0.90	36014
0.95	41166