PENSIONS AND GENETICS: CAN LONGEVITY GENES BE RELIABLE RISK FACTORS FOR ANNUITY PRICING?

By Angus Macdonald and Kenneth McIvor

ABSTRACT

We consider a number of gene variants that have been found to affect longevity. Their effects have been modelled using Cox or logistic regressions, whose fitted parameters have simple asymptotic sampling distributions. The expected present value of a life annuity allowing for these genetic risk estimates inherits a sampling distribution, which can be found by simulation. We find that possibly significant uncertainty about annuity premiums may be overlooked if the standard errors of parameters estimated in medical studies are ignored by medical underwriters. Such considerations may play an important part when the acceptability of using a risk factor in underwriting is conditional on proof of its relevance and reliability. This is the current position in respect of genetic information in many countries, most prominently in the UK.

KEYWORDS

Annuities; Genetics; Longevity; Cox Model; Relative Risk; Odds Ratio; APOE

CONTACT ADDRESS

Angus Macdonald, Department of Actuarial Mathematics and Statistics, and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3209; Fax: +44(0)131-451-3249; E-mail: A.S.Macdonald@ma.hw.ac.uk

1. Pension Annuities and Genetics

1.1 *Longevity Genes and Annuity Pricing*

Previous studies of genetics and insurance have focused on genes leading to early onset of disease, affecting life and critical illness insurance. Here, we consider genes affecting the longevity of older people, hence annuity business. Genes associsted with longer lifetimes are called 'longevity genes', while genes associated with shorter lifetimes are called 'frailty genes'. Table 1 lists some of the genes that have been linked to longevity. Most of them are related to Alzheimer's disease (AD) and heart disorders.

Medical risks have long been taken into account in pricing insurance, and are increasingly being used in pricing annuities. The extent to which genetic information may be taken into account alongside other medical evidence has been strictly limited in several countries, but in at least one case (the United Kingdom) its use depends on being able to demonstrate the relevance—technical, clinical and actuarial—of a given genetic test (see Macdonald (2007)). Any such demonstration will depend on published genetic epidemiology. The actuarial problem is to extract, from this literature, a basis for a pricing model, including an assessment of its reliability. This work faces two major problems:

Table 1: Genes, and their possible related disorders, that have been repeatedly studied for associations with longevity and have shown significant correlations (De Benedictis *et al.*, 2001).

Gene	Disease
ApoE	Alzheimer's disease, Cardiovascular disease
ApoB	Coronary artery disease
ApoA-IV	Alzheimer's disease
ACE	Myocardial infarction, Cerebral infarction, Alzheimer's disease, Essential hypertension
CYP2D6	Parkinson disease
HLA1 & HLA2	Immune Disorders
P53	Cancer
Factors V, VII	Myocardial infarction
Fibrinogen	Coronary artery disease
Prothrombin	Myocardial infarction
MTHFR	Cardiovascular disease, Cancer
mtDNA	Coronary artery disease, Diabetes, Parkinson disease, Alzheimer's disease
PARP	Unknown

(a) Actuarial questions require age-related rates of disease onset, while many medical questions can be answered by simpler statistics. Thus Macdonald & Pritchard (2000) trawled the large literature on AD (up to about 1998), and found just one study that reported age-related risks in enough detail.
(b) Premium rates based on an epidemiological study are functions of the data underlying that study. Assessing their reliability, in any statistical sense, requires either: (i) access to the data; or (ii) publication of ancillary information, such as the full correlation matrix of any parametric models fitted in the study. Both are quite rare. Lu, Macdonald & Waters (2006) were able to use summary data available from some non-parametric studies of polycystic kidney disease, but the opportunity was unusual.

We focus on three epidemiological studies of potential longevity genes, each of which gives sufficient information to allow us to estimate sampling distributions of annuity rates.
(a) Tan *et al.* (2001) fitted a Cox proportional hazards model using a population of 961 Italians of whom 212 (22%) were centenarians. Twelve genes were considered as risk factors (see Table 2). The fitted relative risks and their sampling variances can be used to bootstrap sampling distributions of annuity premium rates.
(b) Arking *et al.* (2005) studied the KLOTHO genotype in a US Ashkenazi Jewish population with 216 subjects. Denoting one allele (gene variant) F and the other V, the possible genotypes are FF, FV, and VV. Again, a Cox model was fitted, parameterised by relative risk with respect to the FV genotype.
(c) Hayden *at al.* (2005) studied the Apoliprotein E (APOE) genotype in a large cohort of subjects over age 65. This gene has three major alleles, $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$, hence six genotypes: $\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 3$, $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$. They fitted a logistic model

Table 2: List of genes studied in Tan *et al.* (2001) labelled $g = 1, 2, \ldots, 12$.

g	Gene	Sample Size	g	Gene	Sample Size
1	Apob35	787	7	INS−	438
2	Apob39	787	8	INS+	438
3	THO7	555	9	mtDNAhapl-J	547
4	THO8	555	10	mtDNAhapl-U	547
5	THO10	555	11	mtDNAstr-136	393
6	SOD2-T	354	12	mtDNAstr-138	393

(similar to a Cox model) parameterised by relative odds with respect to the most common genotype, $\epsilon 3/\epsilon 3$.

In Section 2 we discuss how to simulate sampling distributions of annuity premiums based on Cox and logistic models. In Section 3 we show selected results. We discuss the implications in Section 4

2. Methods

2.1 *The Cox Model*

The Cox model is a semi-parametric multiplicative hazard regression model. Let t be a suitable timescale (such as age). Individual i ($i = 1, 2, \ldots, n$) has force of mortality $\lambda^i(t)$ of the form:

$$\lambda^i(t; \boldsymbol{Z}_i) = \lambda^0(t) \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \tag{1}$$

where: $\boldsymbol{Z}_i$ is a p-dimensional vector of covariates (risk factors) for individual i; $\lambda^0(t)$ is the baseline force of mortality; and $\boldsymbol{\beta}$ is the p-dimensional vector of regression coefficients. Usually $\boldsymbol{\beta}$ is estimated, and sometimes the baseline hazard $\lambda^0(t)$ as well. Here, $\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)$ defines the relative risk, denoted RR_i for brevity.

The logistic regression model is of the form:

$$\frac{\lambda^i(t; \boldsymbol{Z}_i)}{1 - \lambda^i(t; \boldsymbol{Z}_i)} = \frac{\lambda^0(t)}{1 - \lambda^0(t)} \ \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \tag{2}$$

in which $\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)$ defines the relative odds. Denoting the latter RO_i for brevity, constant relative odds imply non-constant relative risks as follows:

$$RR_i(t) = \frac{\lambda^i(t)}{\lambda^0(t)} = \frac{1}{\lambda^0(t)} \times \frac{\frac{\lambda^0(t)}{1-\lambda^0(t)} \ RO_i}{1 + \frac{\lambda^0(t)}{1-\lambda^0(t)} \ RO_i} = \frac{RO_i}{1 - \lambda^0(t) + \lambda^0(t) \ RO_i}. \tag{3}$$

Andersen *et al.* (1993) is a definitive reference on hazard regression models.

2.2 *Baseline Hazards*

Tan *et al.* (2001) used Italian population mortality statistics from 1994 as an initial estimate of the baseline hazard, but their iterative estimation technique resulted in a

different baseline hazard for males and females combined. A single baseline hazard was required because they wished to model gene × sex interactions. They did not publish their baseline hazard, so as a proxy, we used Italian male and female population mortality in 1994 life tables available online to obtain a baseline hazard rate as follows:

$$\lambda^0(t) = \frac{l_f(t)\ \lambda_f(t) + l_m(t)\ \lambda_m(t)}{l_f(t) + l_m(t)} \tag{4}$$

where $l_f(t)$ and $l_m(t)$ are the standard life table functions for the expected numbers alive at time t, for females and males respectively, assuming that $l_f(0) = l_m(0)$.

For both Arking *et al.* (2005) and Hayden *et al.* (2005) we used as a baseline hazard US life table data from 2000.

2.3 *Sampling Distributions of Relative Risks and Premiums*

The three studies provided slightly different parameter estimates, in all cases with estimated variances or standard deviations.

(a) Tan *et al.* (2001) modelled each of the twelve genes separately, not all twelve simultaneously. For each gene g, they gave a relative risk estimate $\widehat{RR}_g$ for females, plus a gene × sex term $\widehat{RR}_g^{g\times s}$, such that the relative risk for males was $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$.
(b) For each KLOTHO genotype g, Arking *et al.* (2005) gave the regression coefficient β_g (trivially equal to 1 for genotype FV), not distinguishing between males and females.
(c) For each APOE genotype g, Hayden *et al.* (2005) gave relative odds estimates $\widehat{RO}_g$ for males and females.

Given a baseline hazard $\lambda^0(t)$, and a relative risk estimate $\widehat{RR}_g(t)$ (either constant, or from Equation (3)), it is simple to calculate the single premium for a whole-life annuity of 1 per year, payable continuously; denote this $\widehat{P}_g$. The notation emphasises that if we knew the true relative risk $RR_g(t)$ we could compute the true premium rate denoted P_g, but in practice we only obtain the point estimate $\widehat{P}_g$. We could express the premium rate as a function of the relative risk: $P_g = f(RR_g(t))$ and through this relationship $\widehat{P}_g$ inherits a sampling distribution from that of $\widehat{RR}_g$ or $\widehat{RO}_g$. This is our real target of study. The simplest way to find it, given that $f()$ is a somewhat complicated function, is by simulating from the sampling distribution of $\widehat{RR}_g$ or $\widehat{RO}_g$.

Assuming the estimated parameter $\hat{\beta}_g$ for gene or genotype g to be Normal (justified asymptotically), $\exp(\hat{\beta}_g)$ is log-normal, with parameters μ_g and σ_g say. Hence, given $\widehat{RR}_g$ (in the Cox model) or $\widehat{RO}_g$ (in the logistic model) and its estimated standard deviation $S[\widehat{RR}_g]$ ($S[\widehat{RO}_g]$) we can find μ_g and σ_g by equating first and second moments.

Then, repeated sampling from the approximate log-normal distribution will give us a simulated sample from the sampling distribution of $f(RR_g(t))$, the premium rate, as required. In all cases we used 10,000 samples.

3. Results

3.1 *Premiums for Females Based on Tan* et al. *(2001)*

The log-normal densities of $\widehat{RR}_g$ for each gene g for females are shown in the left-hand panels of Figures 1 and 2. For each of 10,000 samples from each distribution, we calculated whole-life annuity premiums, for a female age 60 and force of interest $\delta = 0.05$ per annum. The simulated sampling densities of these premiums (as a proportion of the premium with $RR = 1$) are shown in the right-hand panels of Figures 1 and 2.

The most dispersed sampling (premium) distribution is that for a carrier of the INS− gene ($g = 7$). This has a small sample size, 438, compared with (for example) Apob35, with a sample size of 787. However the most uncertain premium estimates exist for those genes that have an extreme relative risk estimate. For example, mtDNAstr-138 has the smallest $\widehat{RR}_g$, 0.275, and its standard deviation is about average, but it produces a very dispersed premium estimate. In other words, it is not necessarily $S[\widehat{RR}_g]$ that dictates $S[\widehat{P}_g]$, but also the magnitude of $\widehat{RR}_g$.

Table 3 presents some key statistics of the premium sampling distributions. In particular, we show some percentiles because they might have a rôle in deciding when a test for a particular genotype might be regarded as a relevant and reliable indicator of increased risk, given the available studies. We discuss such criteria briefly in Section 4.

3.2 *Premiums for Males Based on Tan* et al. *(2001)*

Tan *et al.* (2001) defined relative risks for males with respect to females, equivalent to introducing a gene × sex interaction term. For example, suppose gene g halves female risk ($\widehat{RR}_g = 0.5$) but has no effect on males. Then the interaction term, denoted $\widehat{RR}_g^{g\times s}$, would be 2, so that overall the relative risk for males is $\widehat{RR}_g \times \widehat{RR}_g^{g\times s} = 0.5 \times 2 = 1$.

If we assume: (a) that the sampling distributions of the $\widehat{RR}_g^{g\times s}$ are also log-normal; and (b) that the estimates $\widehat{RR}_g$ and $\widehat{RR}_g^{g\times s}$ are independent, given the data, then we can easily find the sampling distributions of $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$, since the product of log-normal(μ_1, σ_1^2) and log-normal(μ_2, σ_2^2) random variables is log-normal($\mu_1 + \mu_2$, $\sigma_1^2 + \sigma_2^2$). Assumption (a) may be reasonable, but independence is very unlikely, since given the sampling distribution of the overall relative risk for males, any shift in the marginal sampling distribution of $\widehat{RR}_g$ is likely to be compensated for by an opposite shift in the marginal sampling distribution of $\widehat{RR}_g^{g\times s}$. However, there is nothing we can do about this, lacking the sampling covariances of $\widehat{RR}_g$ and $\widehat{RR}_g^{g\times s}$. We can only comment that the results for males may overstate the sampling variances and thus are more tentative than the results for females.

In Figures 3 and 4 we show the log-normal sampling distributions of the relative risk estimates in the left-hand panels, alongside the simulated densities of the single premium for a male age 60 based on force of interest $\delta = 0.05$ per annum. The premium sampling distributions are generally more dispersed than those for females. This is consistent with the fact that there were fewer male centenarians in the study. Some key statistics are shown in Table 3.

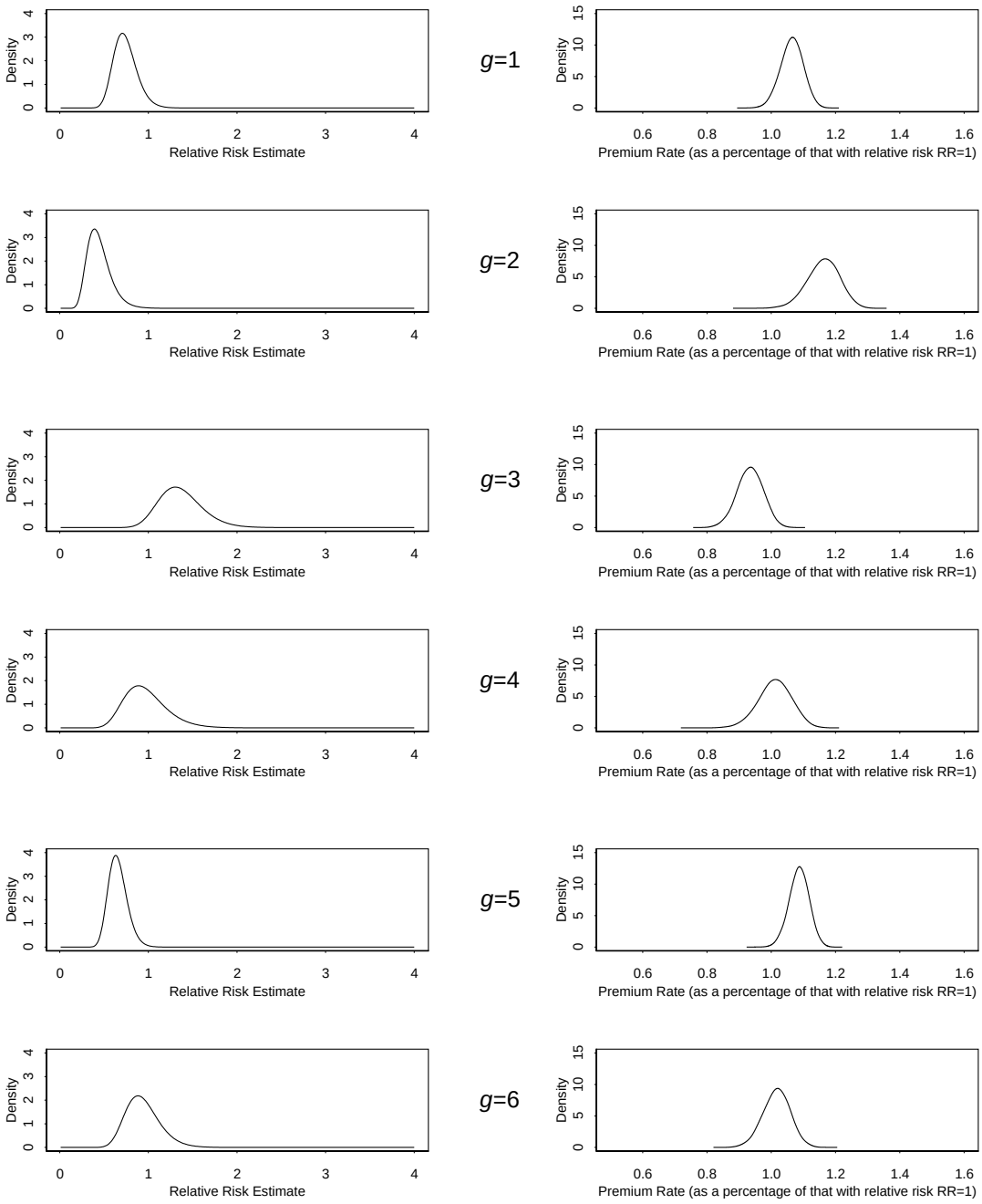


Figure 1: The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 1, \ldots, 6$.

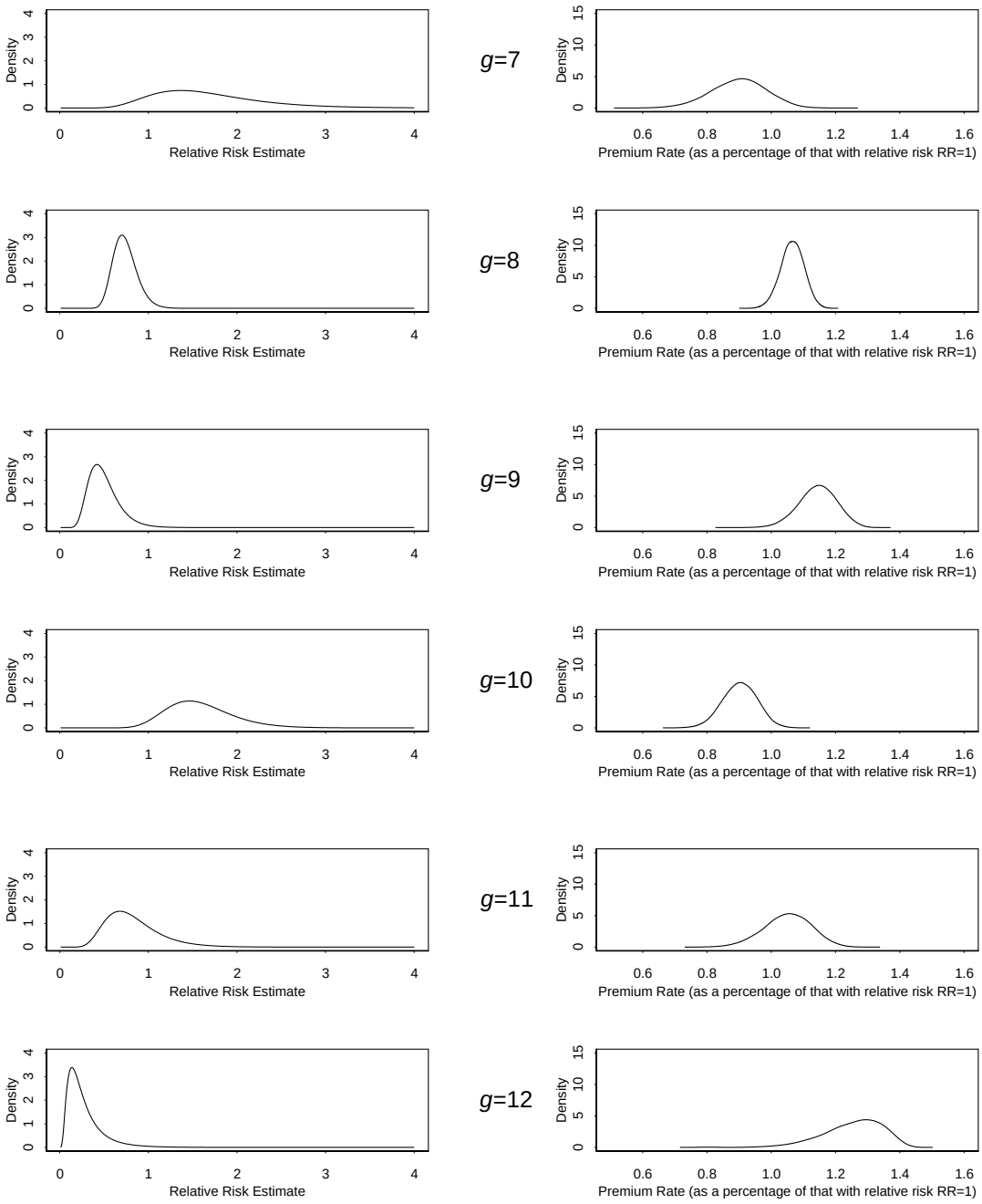


Figure 2: The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 7, \ldots, 12$.

Table 3: The mean, standard deviation and quantiles of single premiums for a whole-life annuity for a female age 60 based on a log-normal distribution of relative risk estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$.

				Quantiles of the Premium Distribution as a Percentage of the Baseline Premium								
Sex	Gene	Mean	St. Dev.	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
		%	%	%	%	%	%	%	%	%	%	%
Female	Apob35	106.4	3.5	99.4	100.6	101.9	104.1	106.5	108.8	110.8	111.9	113.0
	Apob39	116.2	4.9	106.0	107.7	109.8	113.0	116.4	119.7	122.3	124.0	125.2
	THO7	93.4	4.0	85.3	86.7	88.3	90.8	93.5	96.2	98.6	99.9	101.0
	THO8	101.1	5.1	90.8	92.5	94.6	97.9	101.3	104.6	107.5	109.2	110.6
	THO10	108.6	3.1	102.5	103.4	104.7	106.6	108.7	110.7	112.5	113.5	114.5
	SOD2-T	101.6	4.2	93.2	94.6	96.1	98.9	101.7	104.5	106.9	108.3	109.6
	INS−	89.9	8.4	72.7	75.6	79.0	84.2	90.2	95.7	100.5	103.3	105.4
	INS+	106.5	3.6	99.4	100.5	101.9	104.2	106.6	109.0	111.1	112.3	113.2
	mtDNAhapl-J	114.4	5.8	102.6	104.5	106.9	110.7	114.6	118.5	121.7	123.6	125.1
	mtDNAhapl-U	90.2	5.3	79.7	81.5	83.3	86.6	90.3	93.9	97.0	98.6	100.3
	mtDNAstr-136	105.0	7.3	89.8	92.5	95.3	100.3	105.3	110.1	114.1	116.5	118.5
	mtDNAstr-138	125.3	9.4	104.2	108.3	112.7	119.8	126.7	132.3	136.1	137.9	139.3
Male	Apob35	106.8	6.5	93.4	95.5	98.4	102.5	107.1	111.4	115.0	117.0	119.0
	Apob39	105.2	14.7	73.6	79.1	85.6	96.0	106.5	115.9	123.4	127.5	130.0
	THO7	99.3	6.2	86.6	88.8	91.3	95.3	99.5	103.6	107.1	109.2	110.9
	THO8	110.2	5.3	99.4	101.1	103.2	106.8	110.4	113.9	116.8	118.5	120.0
	THO10	103.1	5.9	91.1	93.1	95.3	99.1	103.3	107.3	110.5	112.3	113.9
	SOD2-T	99.6	7.8	83.4	86.6	89.5	94.5	99.8	105.1	109.5	112.1	114.1
	INS−	104.0	14.7	71.7	77.6	84.3	94.6	105.3	114.7	122.1	125.8	128.7
	INS+	99.4	7.6	84.3	86.7	89.4	94.3	99.5	104.6	109.0	111.4	113.6
	mtDNAhapl-J	117.5	5.1	106.9	108.6	110.7	114.3	117.8	121.0	124.0	125.6	127.0
	mtDNAhapl-U	86.2	11.6	62.4	66.3	71.2	78.5	86.5	94.4	101.2	104.8	107.8
	mtDNAstr-136	106.3	9.3	86.9	90.3	94.2	100.2	106.6	112.8	117.9	120.7	123.1
	mtDNAstr-138	125.0	8.0	107.1	110.8	114.5	120.1	125.9	130.8	134.4	136.1	137.6

3.3 *Premiums Based on Arking* et al. *(2005)*

The sampling distributions of the relative risks and (by simulation) of annuity premiums are shown in Figure 5. For brevity we omit the table of statistics. Both KLOTHO genotypes are detrimental to survival (relative to the FV genotype) and therefore significantly reduce premium rate estimates. As we would expect from such a small study (216 participants), the sampling distributions are very dispersed. However, these relative risks have to be interpreted with caution. Being based on individuals over age 95, it is possible that the detrimental effects may be limited to very elderly populations.

3.4 *Premiums Based on Hayden* et al. *(2005)*

Figure 6 shows the sampling distributions of premium rates (for a whole-life annuity issued to a life aged 65, with force of interest $\delta = 0.05$) for APOE genotypes $\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$, relative to the premium rates in respect of the $\epsilon 3/\epsilon 3$ genotype. These are based on 10,000 simulations from a log-normal sampling distribution of the relative odds, using as baseline hazard rates US mortality from calendar year 2000. That is, we attribute population mortality to carriers of the most common genotype, $\epsilon 3/\epsilon 3$.

The notable frailty genotypes seem to be $\epsilon 2/\epsilon 2$ and $\epsilon 2/\epsilon 3$ for females, and $\epsilon 4/\epsilon 4$ for males. The genotype $\epsilon 3/\epsilon 4$ appears to be a frailty genotype in females but a longevity genotype in males, while the opposite is the case for $\epsilon 2/\epsilon 4$.

4. Discussion and Conclusions

4.1 *Acceptable Uncertainty*

Relative risk or relative odds estimates are often published in epidemiological studies. We have highlighted their sampling properties and the sampling distributions inherited by premium rates based on them. Many of the genes in this study might at first sight appear to be financially important; however the sampling distributions of the corresponding premium rates introduce much more uncertainty. This kind of statistical information is relevant to any consideration of using genotype information in insurance practice, for example in the deliberations of the Genetics and Insurance Committee (GAIC) in the UK.

GAIC was charged, by the UK Government, with ensuring that any use of genetic test results by insurers would have a sound actuarial and scientific basis. To date, GAIC has approved only one genetic test (for Huntington's disease, in the case of life insurance). GAIC will face difficult questions if it is required to review tests for more complex disorders, whose results do not indicate an overwhelming increase or decrease in population mortality, and as such sampling error should be taken into account. This is the case for longevity genes. When presented with sampling distributions of relative premium rates, GAIC will have to answer questions such as: what percentile of the premium rate sampling distribution might justify the use of a test by insurers? (we call this an 'acceptance percentile', but it is only one of many criteria that could be adopted). Such questions may well arise as research into genes with modest effects on mortality and morbidity, by current standards, enters the medical mainstream.

If an acceptance percentile were adopted as a criterion for the use of a genetic test, the next question is: what premium loading should be applied? This question is probably

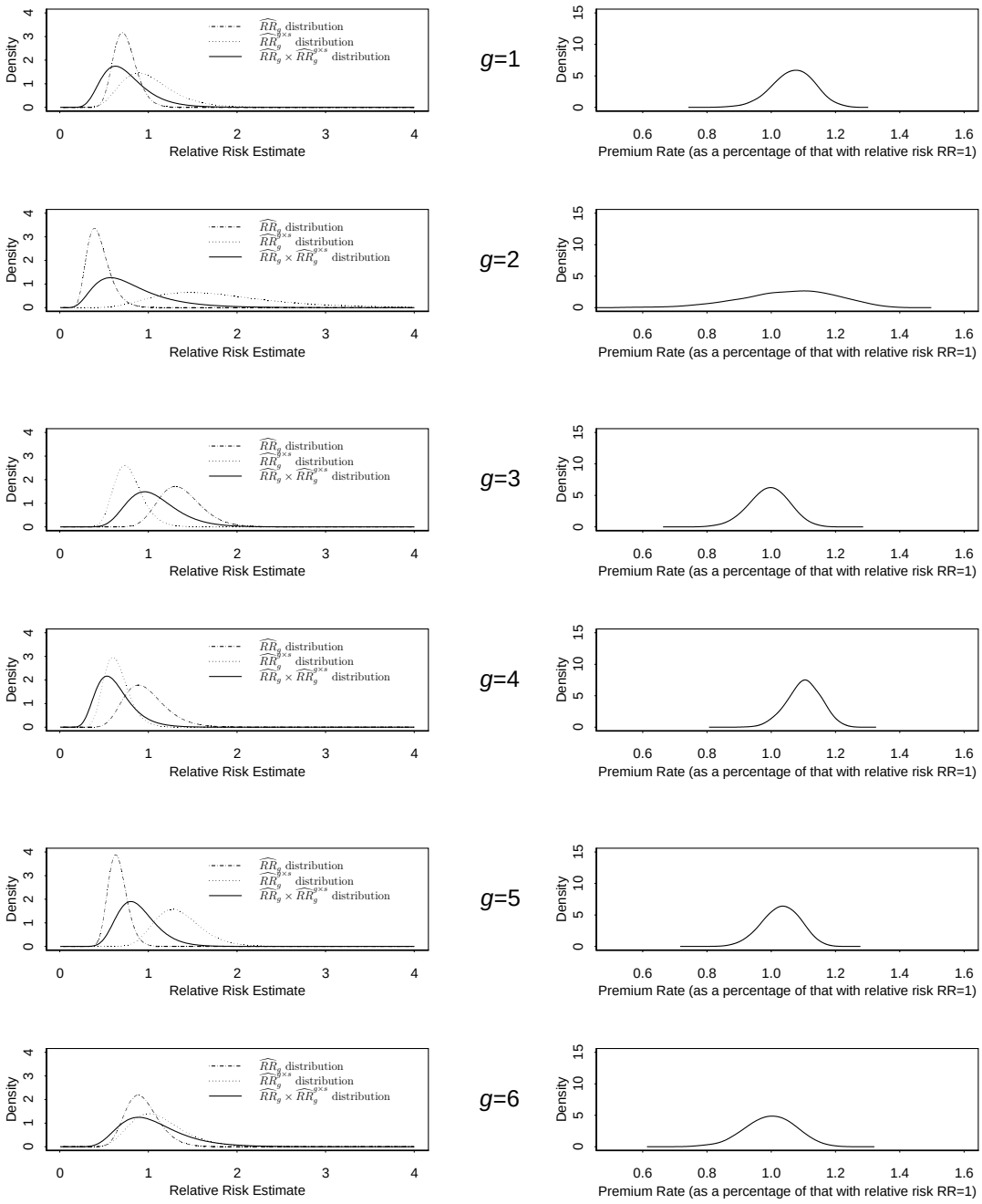


Figure 3: The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 1, \ldots, 6$.

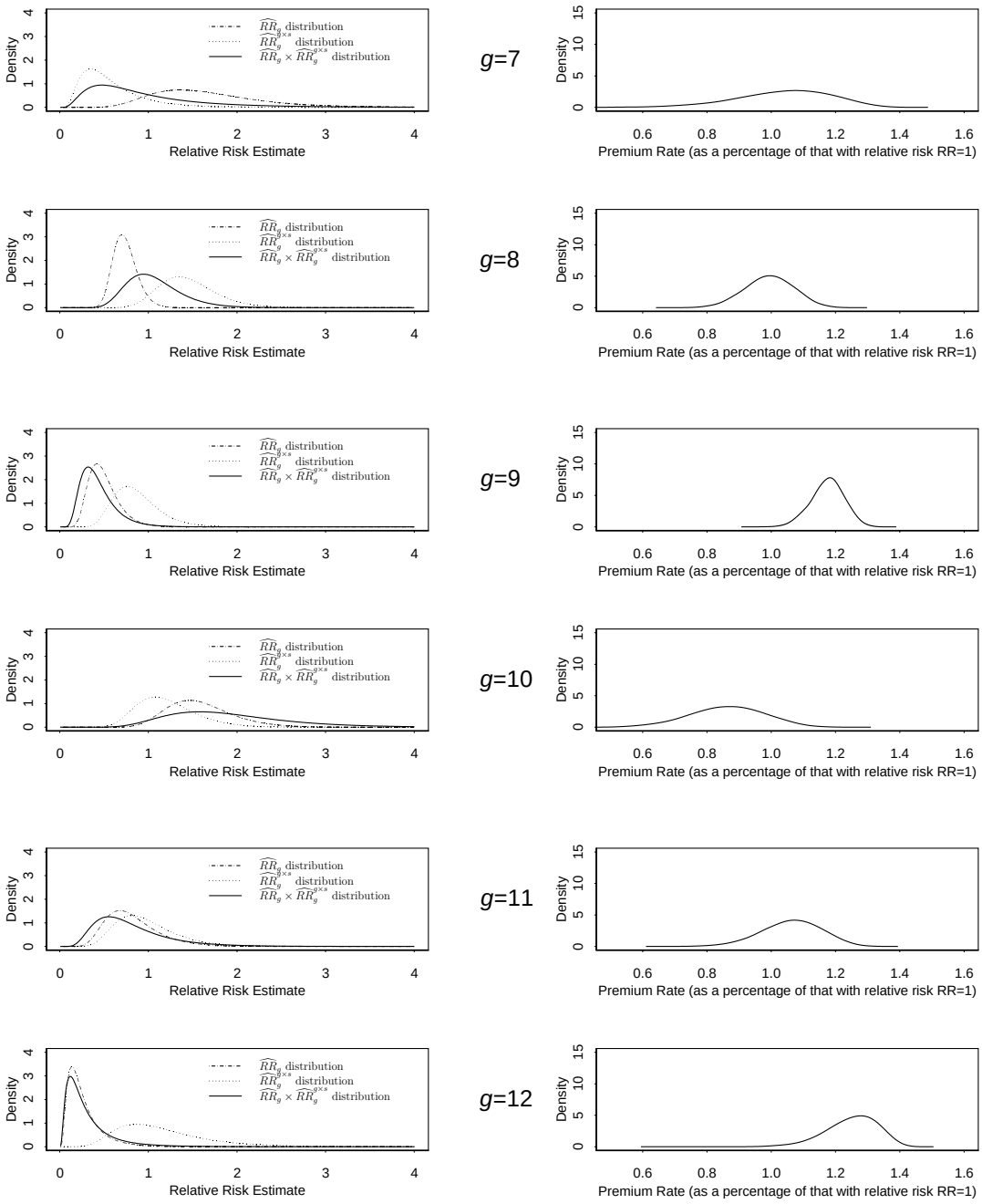


Figure 4: The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 7, \ldots, 12$.

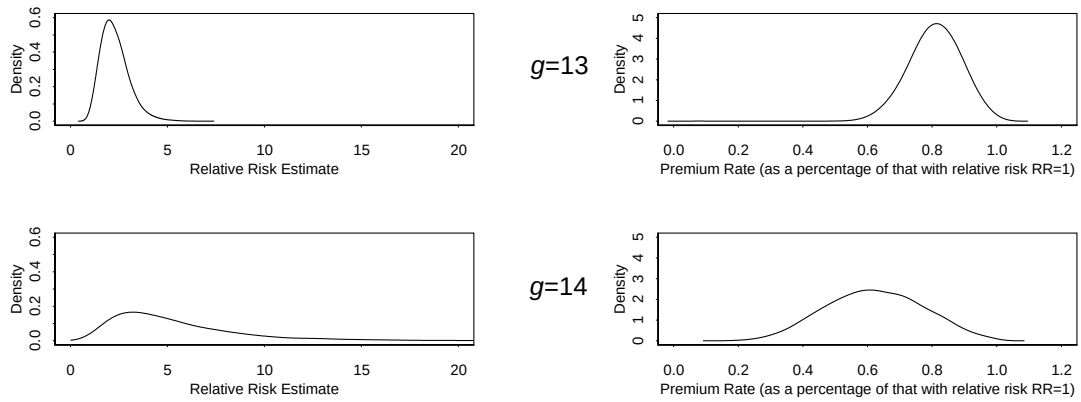


Figure 5: The density curves of log-normally distributed relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for carriers of KLOTHO FF and VV genotypes.

beyond GAIC's remit, and would be left to individual insurers, who would be allowed to take into consideration their different risk tolerances and underwriting practices.

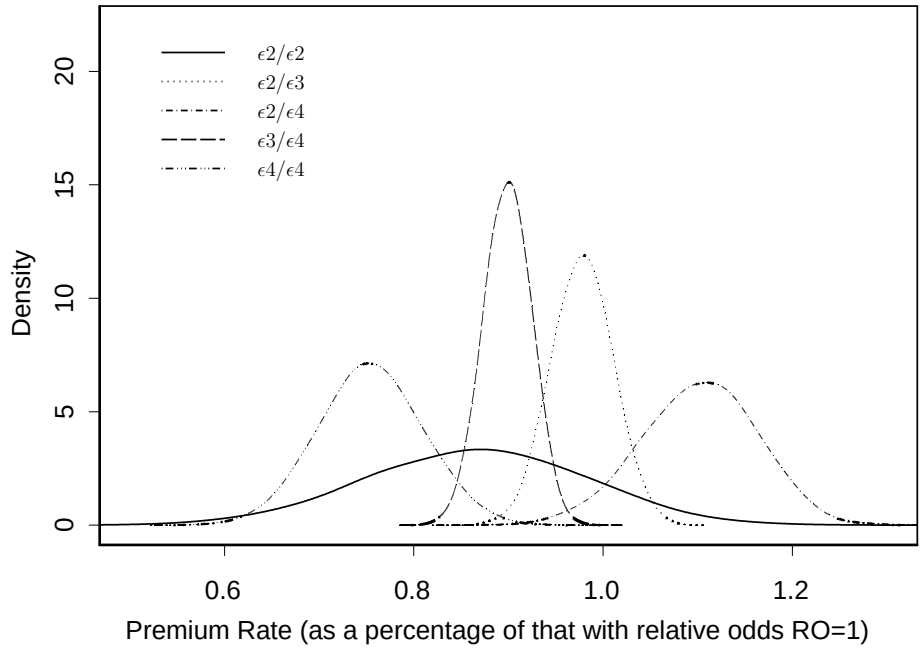
4.2 *Acceptance Percentiles*

As an example of how 'acceptance percentiles' might contribute to such decisions as described above, Table 4 shows which genotypes might be regarded as having a significant impact at a 75%, 90%, 95% and 97.5% level, based on percentiles from Tables 3 and the equivalent tables for KLOTHO and APOE genotypes (not shown). Note that upper or lower percentiles are used, depending on whether a gene is a candidate longevity gene or a candidate frailty gene.

If the criterion of a one-tailed 97.5% confidence interval (of annuity prices) were adopted (implying very low uncertainty) then, among the genes considered by Tan *et al.* (2001), for females, four 'longevity' gene variants would appear to be important: Apob39, THO10, mtDNAhapl-J and mtDNAstr-138. For males, there would be only two: mtDNAhapl-J and mtDNAstr-138. Tan *et al.* (2001) estimated the frequency of these genes in the Italian population: Apob39 and THO10 are both common (frequency $\approx$ 30–40%) whereas mtDNAhapl-J and mtDNAstr-138 are relatively rare (frequency $\approx$ 1–5%). Therefore, in the scenario of widespread genetic testing the genes Apob39 and THO10 could lead to large-scale segmentation of the annuity market (always supposing that results based on an Italian population generalise to other populations).

The APOE genotype is arguably more important. Commercial testing for APOE genotype is readily available and Hayden *et al.* (2005) is only one of many research teams that have confirmed its rôle in longevity. Most APOE genotypes are frailty genotypes that act to reduce annuity premiums (relative to the $\epsilon 3/\epsilon 3$ norm).

Our methodology is not confined to genetic risk. Indeed we are surprised that it has taken genetic risk to draw attention to sampling issues in actuarial estimates based on epidemiological and medical studies. Consideration of premium rate sampling error would seem to be an elementary extension from professional statistical practice to professional



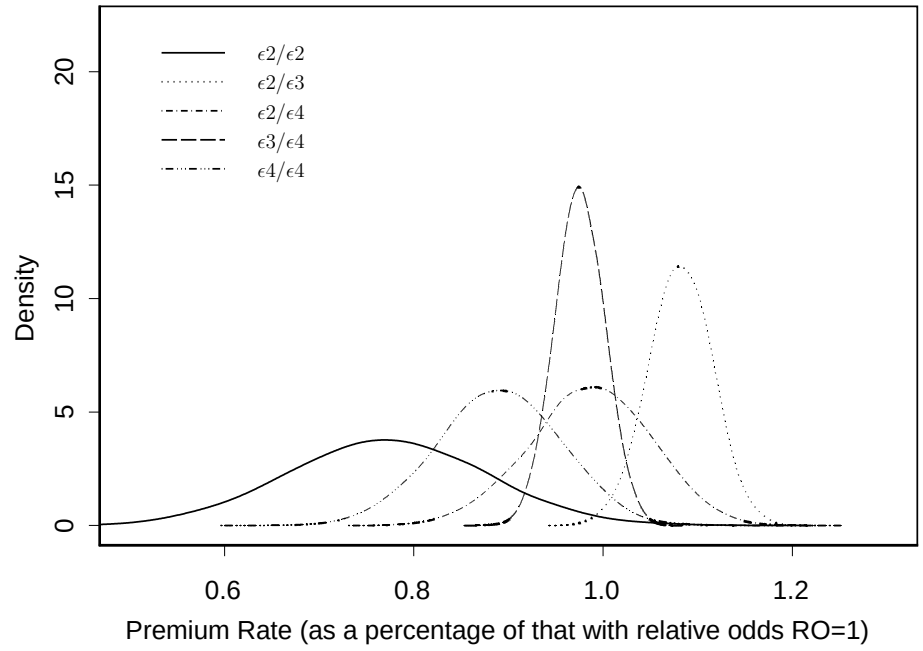


Figure 6: The empirical densities of whole-life annuities for a female (top) and a male (bottom) beginning at age 65, for APOE genotypes $\epsilon2/\epsilon2$, $\epsilon2/\epsilon3$, $\epsilon2/\epsilon4$, $\epsilon3/\epsilon4$, and $\epsilon4/\epsilon4$ relative to the annuity cost of a $\epsilon3/\epsilon3$ genotype carrier.

Table 4: A list of all genes/genotypes studied, and whether they are significant at a 75%, 90%, 95% or 97.5% level. A ✓ represents a significant gene/genotype and a ✗ represents a non-significant gene/genotype. The phenotype is the observable manifestation of the gene/genotype, this is either frailty or longevity.

Gender	Gene/Genotype	75%	90%	95%	97.5%	Phenotype
Female	Apob35	✓	✓	✓	✗	Longevity
	Apob39	✓	✓	✓	✓	Longevity
	THO7	✓	✓	✓	✗	Frailty
	THO8	✗	✗	✗	✗	Longevity
	THO10	✓	✓	✓	✓	Longevity
	SOD2-T	✗	✗	✗	✗	Longevity
	INS-	✓	✗	✗	✗	Frailty
	INS+	✓	✓	✓	✗	Longevity
	mtDNAhapl-J	✓	✓	✓	✓	Longevity
	mtDNAhapl-U	✓	✓	✓	✗	Frailty
	mtDNAstr-136	✓	✗	✗	✗	Longevity
	mtDNAstr-138	✓	✓	✓	✓	Longevity
	APOE $\epsilon 4/\epsilon 4$	✓	✓	✓	✓	Frailty
	APOE $\epsilon 3/\epsilon 4$	✓	✓	✓	✓	Frailty
	APOE $\epsilon 2/\epsilon 4$	✗	✗	✗	✗	Longevity
	APOE $\epsilon 2/\epsilon 3$	✓	✗	✗	✗	Frailty
	APOE $\epsilon 2/\epsilon 2$	✓	✓	✓	✗	Frailty
Male	Apob35	✓	✗	✗	✗	Longevity
	Apob39	✗	✗	✗	✗	Longevity
	THO7	✗	✗	✗	✗	Frailty
	THO8	✓	✓	✓	✗	Longevity
	THO10	✗	✗	✗	✗	Longevity
	SOD2-T	✗	✗	✗	✗	Frailty
	INS-	✗	✗	✗	✗	Longevity
	INS+	✗	✗	✗	✗	Frailty
	mtDNAhapl-J	✓	✓	✓	✓	Longevity
	mtDNAhapl-U	✓	✗	✗	✗	Frailty
	mtDNAstr-136	✓	✗	✗	✗	Longevity
	mtDNAstr-138	✓	✓	✓	✓	Longevity
	APOE $\epsilon 4/\epsilon 4$	✓	✗	✗	✗	Frailty
	APOE $\epsilon 3/\epsilon 4$	✗	✗	✗	✗	Longevity
	APOE $\epsilon 2/\epsilon 4$	✓	✓	✗	✗	Frailty
	APOE $\epsilon 2/\epsilon 3$	✓	✓	✓	✓	Frailty
	APOE $\epsilon 2/\epsilon 2$	✓	✓	✓	✓	Frailty
Both	KLOTHO FF	✓	✓	✓	✓	Frailty
	KLOTHO VV	✓	✓	✓	✓	Frailty

actuarial practice.

Acknowledgements

This work was carried out at the Genetics and Insurance Research Centre at Heriot-Watt University. We would like to thank the sponsors for funding, and members of the Steering Committee for helpful comments at various stages. We are grateful to an anonymous referee whose comments greatly improved the paper. KM was funded by the Engineering and Physical Sciences Research Council.

References

Andersen, P.K., Borgan, Ø., Gill, R.D., & Keiding, N. (1993). *Statistical models based on counting processes*. Springer-Verlag, New York.

Arking, D.E., Atzmon, G., Arking, A., Barzilai, N., & Dietz, H.C. (2005). Association between a functional variant of the KLOTHO gene and high-density lipoprotein cholesterol, blood pressure, stroke, and longevity. *Circulation Research*, **96**, 412–418.

De Benedictis, G., Tan, Q., Jeune, B., Christensen, K., Ukraintseva, S.V., Bonafè, M., Franceschi, C., Vaupel, J.W., & Yashin, A.I. (2001). Recent advances in human gene-longevity association studies. *Mecanisms of Ageing and Development*, **122**, 909–920.

Hayden, K.M., Zandi, P.P., Lyketsos, C.G., Tschanz, J.T., Norton, M.C., Khachaturian, A.S., Pieper, C.F., Welsh-Bohmer, K.A., & Breitner, J.C.S. (2005). Apolipoprotein E genotype and mortality: Findings from the Cache County study. *Journal of the American Geriatrics Society*, **53**, 935–942.

Lu, L., Macdonald, M., & Waters, H. (2008). Premium rates based on genetic studies: How reliable are they?. *Insurance: Mathematics & Economics*, **42**, 319–331.

Macdonald, A.S. (2007), Insurance, in *Handbook of Statistical Genetics (3rd edition)*, eds. Balding, D.J., Bishop, M. & Cannings, C.,John Wiley, Chichester.

Macdonald, A.S. & Pritchard, D.J. (2000). A mathematical model of Alzheimer's disease and the ApoE gene. *ASTIN Bulletin*, **30**, 69–110.

Tan, Q., De Benedictis, G., Yashin, A.I., Bonafè, M., DeLuca, M., Valensin, S., Vaupel, J.W., & Franceschi, C. (2001). Measuring the genetic influence in modulating the human life span: gene-environment interaction and the sex-specific genetic effect. *Biogerontology*, **2**, 141–153.