SAMPLING DISTRIBUTIONS OF CRITICAL ILLNESS INSURANCE PREMIUM RATES: BREAST AND OVARIAN CANCER

BY

LI LU, ANGUS MACDONALD AND HOWARD WATERS

ABSTRACT

Evaluating the risk of disorders in long-term insurance often relies on rates of onset estimated from quite small epidemiological studies. These estimates can carry considerable uncertainty, hence so may functions of them, such as a premium rate. In the case of genetic disorders, where it may be required to demonstrate the reliability of genetic information as a risk factor, such uncertainty may be material. Epidemiological studies publish their results in a variety of forms and it is rarely easy to estimate the sampling distribution of a premium rate without access to the original data. We found a large study of breast and ovarian cancer that cited relative risks of breast and ovarian cancer onset, with confidence intervals, in 10-year age groups. We obtained critical illness premium rates and their sampling distributions by parametric bootstrapping, and investigated the effect of possible patterns of sampling correlations. We found that this study provides ample statistical evidence that known BRCA1 or BRCA2 mutations, or a typical family history of breast or ovarian cancer, are reliable risk factors, but the sampling covariances of the relative risks could be important at some ages and terms. Studies that cite only standard errors of parameter estimates erect a small but awkward barrier between the models they describe, and some important actuarial questions.

KEYWORDS

Breast Cancer; Critical Illness Insurance; Family History; Genetics; Ovarian Cancer; Parametric Bootstrapping; Sampling Distributions.

1. INTRODUCTION

1.1. Uncertainty in the Estimation of Premium Rates

Underwriting the risk of many disorders, and genetic disorders in particular, has usually relied on published epidemiological studies. These rarely include the original data, for reasons of confidentiality and brevity. Therefore, while the actuary may use such studies to estimate relevant quantities, like premium

rates, it is rarely possible to quantify the sampling properties of these quantities. Since many medical studies involve small samples, compared with those needed to establish reliable life tables, the sampling uncertainty of premium rates based on them is unknown but could be large. The question we address is: if we regard premium rates as point estimates, how may we estimate their sampling properties, inherited from the data in the underlying medical studies?

In the UK, this question has extra force, because the insurance industry has agreed not to use genetic test results for underwriting unless permitted to do so by a quasi-governmental body called the Genetics and Insurance Committee (GAIC). Even then, they could only be used for very large sums assured (Daykin *et al.*, 2003). One of GAIC's criteria is the reliability with which a genetic test result might identify increased risk. It has not laid down specific methods of measuring such reliability, but it seems reasonable that the sampling distributions of any key parameters estimated from genetic studies should contribute to that assessment.

Lu, Macdonald & Wekwete (2007) studied one approach to obtaining the sampling distribution of premium rates based upon genetic epidemiological studies of modest size. This is the only published study of which we are aware, and it was limited in a number of ways:

(a) They considered adult polycystic kidney disease (APKD), which is a rare and purely genetic disorder. Therefore the existence of a family history identifies the presence of a causative mutation in the family, which leads to relatively straightforward analysis. But other, major, disorders have both genetic and non-genetic causes, such as lifestyle and diet, which must both be modelled.

(b) They applied bootstrapping methods to exposure and event data suitable for non-parametric (Kaplan-Meier or Nelson-Aalen) estimation. However in many genetic studies, parametric models are estimated. Even when the number of parameters is small, it is usual to publish their standard errors, but not the full information matrix (assuming likelihood methods to be used).

We will consider breast cancer (BC) and ovarian cancer (OC), common diseases of which a small proportion can be attributed to specific gene mutations. We use a recent study (Antoniou *et al.*, 2003) which estimated age-related rates of onset using a piecewise constant parametric model. Standard errors but not the the full information matrix were given. Our aim is to approximate the sampling distributions of critical illness (CI) insurance premium rates allowing for familial risk of BC/OC. In doing so, we are led to study the effect of possible patterns of correlation among the parameter estimates.

1.2. Genetics of Breast and Ovarian Cancer

Two germline mutations in BRCA1 and BRCA2 genes have so far been found to confer high risk of BC and OC. BRCA1 mutations account for 2% of all cases of BC, BRCA2 mutations for a smaller proportion. Early studies reported
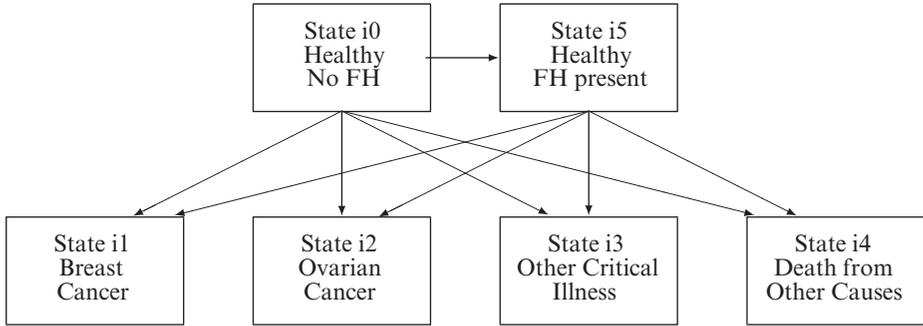
FIGURE 1: A model of breast and ovarian cancer, development of a family history and critical illness insurance, in subpopulation *i* defined by individual and familial genotype.
'FH' = Family History.

the cumulative risk of BC by age 70 to be over 80% (Easton *et al.*, 1993; Easton *et al.*, 1995) for a BRCA1 mutation carrier; a more recent study reduced the risk to 65% (Antoniou *et al.*, 2003). The corresponding risk for a BRCA2 mutation carrier is 45% (Antoniou *et al.*, 2003). The cumulative risk by age 70 of OC is about 39% among BRCA1 mutation carriers and 11% among BRCA2 mutation carriers (Antoniou *et al.*, 2003). Estimates of the prevalence of BRCA1 and BRCA2 mutations in the general population vary from 0.0005 to 0.001 and from 0.0001 to 0.002, respectively. Gui *et al.* (2006) summarised these and chose the estimates of Antoniou *et al.* (2002), namely 0.00058 (BRCA1) and 0.00068 (BRCA2). We will do the same.

1.3. A Dynamic Family History Model

Having a family history of a disease is, in most cases, much more important in insurance practice, than having a genetic test result. This is because it often may be used by underwriters, even if genetic tests may not be. In this context, the difference between purely genetic disorders and other disorders is material.

The distinguishing feature of a purely genetic disease is that one or more inheritable gene mutations cause it and non-mutation carriers cannot develop it. A person is at risk only if they inherited a mutation from either parent, so in many cases it is reasonable to simplify the definition of a family history as follows:

Family History is Present = Affected Parent

see, for example, Lu, Macdonald & Wekwete (2007). However this breaks down in the case of disorders that can have non-genetic causes, including BC and OC. These models must take into account the possibility that a family history could arise by chance, with no mutation being present.

Lu, Macdonald & Waters (2007) and Gui *et al.* (2006) presented such a dynamic model of a family history of BC/OC and other similar disorders. Lu (2006) described in detail how to parameterise the model and how to estimate critical illness (CI) insurance premiums. In the UK, CI insurance usually pays a lump sum on the occurrence of diagnosis of cancer, heart attack, stroke or other serious disorders as specified in the policy. Figure 1 shows a continuous-time Markov model that represents the life history of a woman who applies for CI insurance. We model the onset of BC and OC as transitions into the states labelled $i1$ and $i2$. Some key points about this model are as follows.

(a) The most important probability we have to obtain is:

P[Insurance Applicant Carries a Gene Mutation | Family History Present].

(b) We label genotypes 1 to indicate absence of mutations, 2 to indicate presence of a BRCA1 mutation and 3 to represent presence of a BRCA2 mutation.

(c) We partition the whole population into five subpopulations, labelled $i = 1, 2, 3, 4, 5$, as shown in Table 1. The subdivision determines: (i) the presence (or otherwise) of a mutation in the applicant's family, denoted by f_i; and (ii) the applicant's own genotype, denoted by g_i. The proportion of the population in each subpopulation is derived from the population mutation frequencies (see Gui *et al.* (2006) for details of the calculation).

(d) A woman's risk of BC and OC depends on her own genotype g_i. Conditional on g_i, her risk is independent of her family history.

(e) Given a detailed enough definition of a family history (FH) we can pinpoint its appearance as an event at a specific time, and therefore model it as a transition between states, as shown in Figure 1 (from state 'Healthy, No FH' to state 'Healthy, FH Present'). We can calculate the corresponding transition intensity given a model of family structure (see Section 4).

TABLE 1

SUBPOPULATIONS DEFINED BY AN INSURANCE APPLICANT'S OWN GENOTYPE AND MUTATIONS PRESENT IN THEIR FAMILY. DATA SOURCE: GUI *ET AL.* (2006).

	Mutations		Genotypes		Proportion of Population
Subpopulation	Present in Family	Carried by Applicant	Family's	Applicant's	
$i = 1$	None	None	$f_i = 1$	$g_i = 1$	0.994972360
$i = 2$	BRCA1	None	$f_i = 2$	$g_i = 1$	0.001164102
$i = 3$	BRCA1	BRCA1	$f_i = 2$	$g_i = 2$	0.001164102
$i = 4$	BRCA2	None	$f_i = 3$	$g_i = 1$	0.001349718
$i = 5$	BRCA2	BRCA2	$f_i = 3$	$g_i = 3$	0.001349718

Given estimates of all the intensities in all five subpopulations, we can calculate (using Thiele's equations) the expected present values of any insurance cashflows for: (a) identified BRCA1 or BRCA2 mutation carriers; or (b) an applicant who has 'developed' a family history. In the latter case, EPVs are weighted averages of EPVs given presence in states 15, 25, 35, 45 and 55, the weights being the occupancy probabilities in each of these states at the applicant's age.

The rate of onset of 'other critical illnesses' is based on the model described in Gutiérrez & Macdonald (2003). The rate of mortality is based on the English Life Tables No.15, adjusted to exclude deaths which follow a CI insurance claim. We refer the reader to Gutiérrez & Macdonald (2003) for details.

We used a force of interest of 0.05 per annum and a fourth-order Runge-Kutta method with step size 0.0005 years to solve Thiele's equations and the Kolmogorov forward equations, as needed.

1.4. More on Sampling Uncertainty in Actuarial Models

Consider a continuous-time Markov model with state space $S = \{0, 1, 2, \ldots, n\}$. In a classic actuarial framework, an insurance contract is specified by continuous payments at rate b_x^j per unit time at age x while staying in state j, and a sum assured b_x^{jk} paid on transition at age x from state j to state k. Given the force of interest δ, the present value (PV) of payments is a random variable Y. The distribution of Y is determined by the set of transition intensities $\mu = \{\mu_x^{jk} : j, k \in S\}$. Y also depends on the payment functions b_x^j and b_x^{jk} but in this study we assume they are known, and consider only the uncertainty introduced by having to estimate μ. We are interested in the sampling distribution of the expected present value (EPV) of the payments, denoted by $\mathrm{E}_\mu(Y)$.

(a) Let $\tilde{\mu}$ denote the estimator of μ. Once we obtain a point estimate $\hat{\mu}$, given some data, we usually estimate $\mathrm{E}_\mu(Y)$ by $\mathrm{E}_{\hat{\mu}}(Y)$.

(b) The point estimate $\hat{\mu}$ can be regarded as a single sample drawn from the distribution of the estimator $\tilde{\mu}$, denoted by F.

(c) Following Lu, Macdonald & Wekwete (2007) we are interested in $\mathrm{Var}_F[\mathrm{E}_{\tilde{\mu}}(Y)]$ as a measure of the uncertainty about the EPV, induced by the finite sample size. More generally, we would like to estimate F itself.

(d) We are interested mainly in the intensities μ_x^{i01} and μ_x^{i02} in Figure 1, the rates of onset of BC and OC, and in the intensity of onset of family history, which will depend on μ_x^{i01} and μ_x^{i02}. We fix the other intensities at their estimated values from Gutiérrez & Macdonald (2003). We denote this set of intensities, some fixed and others to be estimated, $\mu^\star$, with estimate $\hat{\mu}^\star$ given some data, whose corresponding estimator $\tilde{\mu}^\star$ has sampling distribution denoted by F^*. So we want to estimate $\mathrm{Var}_{F^\star}[\mathrm{E}_{\tilde{\mu}^\star}(Y)]$.

The complexity of most EPVs, as functions of age-dependent intensities, means it is difficult to compute $\mathrm{Var}_{F^\star}[\mathrm{E}_{\tilde{\mu}^\star}(Y)]$ directly, but resampling methods applied to the original data are effective. We take the following approach:

(a) A recent study (Antoniou *et al.*, 2003) estimated age-specific log-relative risks
(log(RR)) of BC and OC onset, in respect of mutation carriers relative to
non-carriers. They assumed the relative risks to be piecewise constant over
10-year age groups. 95% confidence intervals were given, assuming multi-
variate normal sampling distributions of the log(RR). See Section 2 for a
description of the data.

(b) The confidence intervals yield approximate variances, but the covariance
(or information) matrix for the estimates in different age groups was
not given. We explore the effect of various patterns of covariance in Sec-
tion 3.

(c) We resample point estimates of the log(RR) by simulating from their
assumed multivariate Normal sampling distribution. The relative risks are
applied to baseline risks, which are the population rates of onset of BC/OC
in 1973-77 (the relevant period).

(d) Each simulation gives a sample from the distribution F^*, from which we can
compute a value of $E_{\hat{\mu}^*}(Y)$, for any relevant insurance cashflows. A large
number of simulations (10,000) provides the empirical distribution function
of $E_{\tilde{\mu}^*}(Y)$, hence an estimate of $Var_{F^*}[E_{\tilde{\mu}^*}(Y)]$.

(e) Using underwriting classes defined in Section 5.2, we obtain premium rates
given knowledge of family history.

2. The Data

Antoniou *et al.* (2003) combined pedigree data from 22 studies to investigate
the average risk of BC/OC. The study involved 8,139 index patients with female
or male breast cancer or invasive epithelial ovarian cancer. Mutations in the
BRCA1 or BRCA2 genes were carried by 500 patients. The records of these
mutation carriers included date of birth, age at diagnosis of BC/OC and age
at death or age at last observation.

Age-specific BC/OC incidence rates were estimated by maximum likelihood.
These were assumed to be of the form $\lambda(x) = \lambda_0(x) \exp[g(x)]$ at age x, where
$\lambda_0(x)$ is the onset rate in England and Wales in 1973-77, and $\exp[g(x)]$ is the
age-specific relative risk (RR) of BC/OC onset in mutation carriers, with respect
to the whole population. The $\lambda_0(x)$ were estimated using data from the Office of
National Statistics (ONS) and the Human Mortality Database (www.mortality.org).
The relative risks were assumed to be constant over 10-year age groups.

Table 2 shows the relative risks. Those for BRCA2 mutation carriers were
only estimated for ages over 40 years, because there were no OC cases diag-
nosed at lower ages.

We estimated the standard errors of the estimated log-relative risks from
the 95% confidence intervals in Table 2. These are shown in Table 3. Note that,
given the genotype, the probability of developing BC was assumed to be inde-
pendent of the probability of developing OC.

TABLE 2

Point estimates of relative risks of breast cancer (BC) and ovarian cancer (OC) in BRCA1 or BRCA2 mutation carriers as compared to incidences for England and Wales in 1973-77. 95% confidence intervals are shown in brackets.
Source: Antoniou et al. (2003).

Age Group	BRCA 1 Mutation		BRCA2 Mutation	
	BC	OC	BC	OC
20-29	17 (4.2-71)	1.0	19 (4.5-81)	1.0
30-39	33 (23-49)	49 (21-111)	16 (9.3-29)	1.0
40-49	32 (24-43)	68 (42-111)	9.9 (6.1-16)	6.3 (1.4-28)
50-59	18 (11-30)	31 (14-66)	12 (7.4-19)	19 (9.0-41)
60-69	14 (6.3-31)	50 (22-114)	11 (6.3-20)	8.4 (2.2-32)

TABLE 3

Standard errors of the estimated log-relative risks.

Age Group	BRCA1		BRCA2	
	BC	OC	BC	OC
20-29	0.72	–	0.73	–
30-39	0.19	0.42	0.29	–
40-49	0.15	0.25	0.25	0.76
50-59	0.25	0.40	0.24	0.38
60-69	0.41	0.42	0.29	0.68

3. SIMULATING FROM THE SAMPLING DISTRIBUTIONS OF THE RELATIVE RISKS

We wish to simulate from the sampling distributions of the log-relative risks, which are assumed to be multivariate Normal. Given estimates of the (vector) mean and the full covariance matrix, this can be done by standard methods, see for example Glasserman (2004). (We used Cholesky decomposition of the covariance matrix.) The problem is that we know only the diagonal entries of the covariance matrix. The off-diagonal sampling covariances are unknown.

The first approximation, in these circumstances, might be to assume the covariances to be zero, hence uncorrelated estimates in each age group. However, we set out to explore the effect that various patterns of covariance might have on the outcomes (the sampling distributions of EPVs).

(a) For our purposes we may as well assume strong rather than weak correlations. Weaker correlations will give results closer to the case of independent estimates.

(b) We use correlation coefficients with values +0.9 and –0.9.

TABLE 4

ASSUMED STRUCTURE OF CORRELATION COEFFICIENTS.
+ INDICATES POSITIVE CORRELATION AND − INDICATES NEGATIVE CORRELATION.

Age Group	S1				S2				S3				S4			
	20	30	40	50	20	30	40	50	20	30	40	50	20	30	40	50
20	1	+	+	+	1	−	+	+	1	+	−	+	1	+	+	−
30	+	1	+	+	−	1	−	−	+	1	−	+	+	1	+	−
40	+	+	1	+	+	−	1	+	−	−	1	−	+	+	1	−
50	+	+	+	1	+	−	+	1	+	+	−	1	−	−	−	1

Age Group	S5				S6				S7				S8			
	20	30	40	50	20	30	40	50	20	30	40	50	20	30	40	50
20	1	−	−	+	1	−	+	−	1	+	−	−	1	−	−	−
30	−	1	+	−	−	1	−	+	+	1	−	−	−	1	+	+
40	−	+	1	−	+	−	1	−	−	−	1	+	−	+	1	+
50	+	−	−	1	−	+	−	1	−	−	+	1	−	+	+	1

(c) A covariance matrix must be positive definite. This limits the possible patterns. Given a 4×4 covariance matrix (since we will not be interested in ages over 60) we identified 8 possible patterns, denoted by S1–S8, of positive and negative correlations. They are shown in Table 4.

(d) There may be more possible patterns but we think these eight give sufficient insight into the effect on the sampling distributions of EPVs.

4. MODELLING ONSET OF FAMILY HISTORY

Following typical insurance practice, we define a 'family history' to mean that two or more first-degree relatives (FDRs, meaning parents and siblings) develop BC or OC before age 50.

Given a person in Subpopulation i we know their own genotype g_i and the genotype f_i present in their family (where $f_i = 1$ means no mutations are present in the family). Let the probability that they have developed a family history by age x be ${}^{i}F_{x}^{fh}$. From this, we can find the intensity μ_{x}^{i05} by numerical differentiation. The following method of estimating ${}^{i}F_{x}^{fh}$ was described in detail in Lu, Macdonald & Waters (2007) and Gui *et al.* (2006). We make the following assumptions:

(a) We consider nuclear families only, in which all siblings are at the same age and the parents are 30 years older. We suppose the mother was healthy at age 30 when she had her children.

TABLE 5

DISTRIBUTION OF THE NUMBER OF THE APPLICANT'S SISTERS.
SOURCE: MACDONALD, WATERS & WEKWETE (2003).

No. of sisters k	Probability $P[k]$	No. of sisters k	Probability $P[k]$
0	0.54759802	4	0.00285702
1	0.33055298	5	0.00035658
2	0.09749316	6	0.00002634
3	0.02111590		

(b) The distribution of the number of the applicant's sisters is taken from Macdonald, Waters & Wekwete (2003) and is shown in Table 5. The probability that she has k sisters is denoted by $P[k]$. We assume the applicant has no more than 6 sisters.

(c) Each of the applicant's sisters in a carrier family (Subpopulations 2-5) is a carrier with probability 1/2.

(d) Only one parent in a carrier family is a mutation carrier. We ignore the remote possibility that both parents are carriers.

We consider a healthy applicant age x in Subpopulation i. Let ${}^i P_x^{0M}$ denote the probability that her mother has not suffered BC and OC by age $x + 30$ or age 50, whichever is lower (including the possibility that she died of something else before reaching that age). Let ${}^i P_x^{0S}$ and ${}^i P_x^{1S}$ denote the probability that no sister, or exactly one sister, respectively, has developed BC or OC by age x. Then ${}^i F_x^{fh}$ can be split into two terms:

$$ {}^i F_x^{fh} = (1 - {}^i P_x^{0M}) \times (1 - {}^i P_x^{0S}) + {}^i P_x^{0M} \times (1 - {}^i P_x^{0S} - {}^i P_x^{1S}). \qquad (1) $$

We can represent the life histories of each of the applicant's relatives by a model similar to that in Figure 1, excluding the 'family history' state. Let $P_{x,t}^{g,BC}$ and $P_{x,t}^{g,OC}$ be the probabilities that a person with genotype g, who is healthy at age x, is in the state representing onset of BC or OC, respectively, at age $x + t$. These can all be computed because we have estimates of all the intensities. Then based on the assumptions above:

$$ {}^i P_x^{0M} = 1 - \frac{1}{2} (P_{30, x \wedge 20}^{fi, BC} + P_{30, x \wedge 20}^{fi, OC} + P_{30, x \wedge 20}^{1, BC} + P_{30, x \wedge 20}^{1, OC}), \qquad (2) $$

where $x \wedge 20$ indicates that onset of BC or OC after age 50 does not contribute to the development of a family history.

Let $P_{50 \wedge x}$ denote the probability that a sister of the applicant has developed BC or OC by age x or by age 50, whichever is less. Then:

$$ P_{x \wedge 50} = \frac{1}{2} (P_{0, x \wedge 50}^{fi, BC} + P_{0, x \wedge 50}^{fi, OC} + P_{0, x \wedge 50}^{1, BC} + P_{0, x \wedge 50}^{1, OC}) \qquad (3) $$

and the other terms in Equation (1) are obtained as:

$$P_x^{0S} = \sum_{k=0}^{6} P[k](1 - P_{x \wedge 50})^k \qquad (4)$$

$$P_x^{1S} = \sum_{k=0}^{6} P[k]k(1 - P_{x \wedge 50})^{k-1} P_{x \wedge 50}. \qquad (5)$$

5. The Sampling Distribution of Critical Illness Insurance Premiums

5.1. The Premium Rate for a Known Mutation Carrier

We consider all the assumed patterns of covariance shown in Table 4, denoted by S1–S8. Each results in a different degree of dispersion of the simulated premium rates. For brevity we choose to show the results obtained under those patterns that achieved the greatest and least degrees of dispersion, together with results assuming independent estimators in respect of each age group.

Figures 2 and 3 illustrate the empirical distributions of the simulated premiums. We express premiums as a percentage of the standard premium rate, which is the premium rate for a known non-carrier. The solid line is the case of no correlation, the dashed line represents the most dispersed result and the dotted line represents the least dispersed result. The correlation patterns yielding these results are indicated.

We observe the following main features:

(a) All-positive correlation (S1 in Table 4) results in the greatest dispersion for almost all entry ages and policy terms. This is not surprising, since the relative risks will tend to be high or low in all age groups, with no offsetting of high risks in some age groups and low risks in others.

(b) No one pattern of correlation accounted for all the least dispersed premium rates. The empirical densities all have a sharp peak and, in most cases, 95% confidence intervals as narrow as ±10% of the simulated mean.

(c) The pattern of correlation makes almost no difference to the means of the simulated premium rates.

(d) We include the results for entry age 50 with policy term 10 years for consistency with Figure 4. However, these are determined by the age group 50-59 years alone, therefore by the marginal distribution of the relative risk estimate for that age group. Hence the results are identical under any pattern of correlation.

5.2. Premium Rates Allowing for a Family History

Consider the premium rate charged to a person age x, who has a family history of BC/OC. We assume insurers may use this information for underwriting,
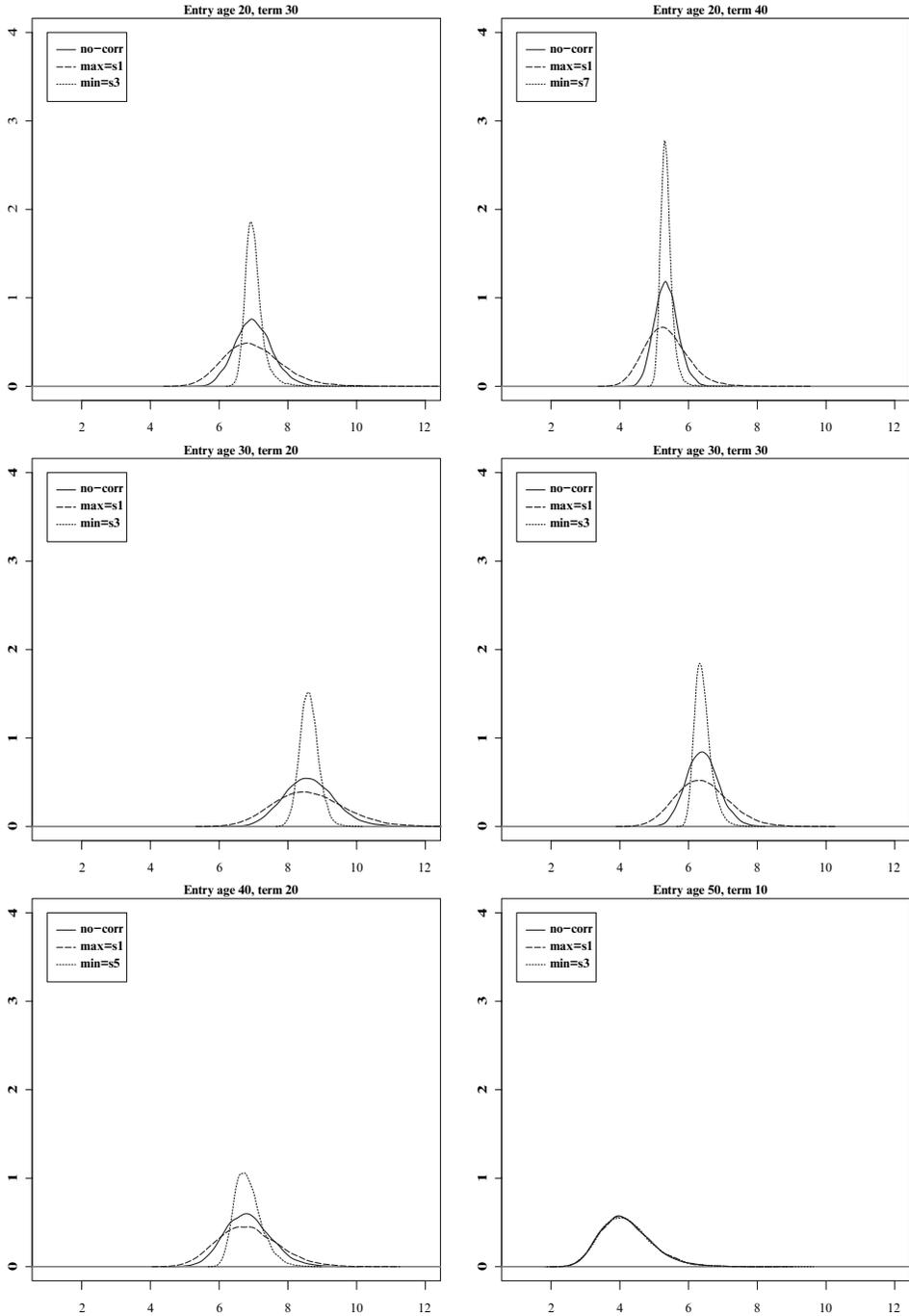
FIGURE 2: Empirical distribution of simulated level premium rates, as a multiple of the standard premium, for a known BRCA1 mutation carrier with various terms and ages at entry. The *x*-axis is the simulated premium rate, as a multiple of the standard premium; the *y*-axis is the density.
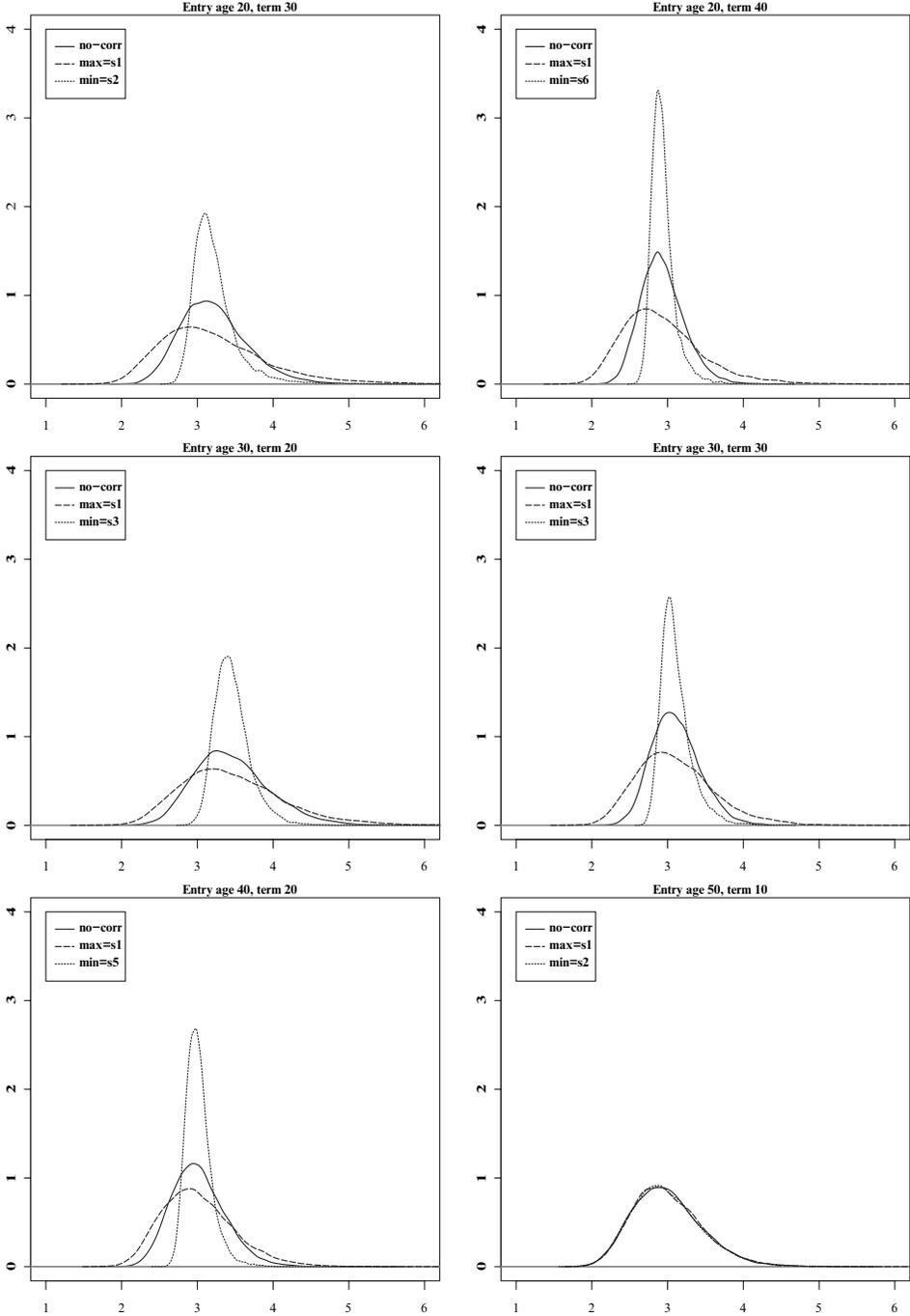
FIGURE 3: Empirical distribution of simulated level premium rates, as a multiple of the standard premium, for a known BRCA2 mutation carrier with various terms and ages at entry. The *x*-axis is the simulated premium rate, as a multiple of the standard premium; the *y*-axis is the density.

which is the practice in the UK but not everywhere else. Hence there are two underwriting classes: the 'ordinary rates' class (denoted by OR) and the 'family history' class (denoted by FH). Let C denote the set of states included in an underwriting class. Let OP_x^{ij} denote the proportion of people in the general population who are in state ij at age x. Let B_x^{ij} and C_x^{ij} be the EPVs of a unit benefit and a unit premium, respectively, conditional on presence in state ij at age x. The EPV of any insurance payments due to a person in an underwriting class is a weighted average of the EPVs in respect of each state in that underwriting class, the weights being the proportions OP_x^{ij}. Therefore the level premium for the underwriting class defined by the set of states C, denoted by LP_x^C, is:

$$ LP_x^C = \frac{\sum_{ij \in C} OP_x^{ij} B_x^{ij}}{\sum_{ij \in C} OP_x^{ij} C_x^{ij}}. \tag{6} $$

We calculated level premiums LP_x^{OR} for the OR class and LP_x^{FH} for the FH class. Figure 4 shows the simulated sampling distributions of the level premium rates LP_x^{FH}, as a percentage of the OR premium rates LP_x^{OR}. We choose the upper bound of the x-axis to be as large as 250% of the OR premium. This choice is consistent with insurance practice, in that most insurers would decline the application if the premium exceeds 250% of the standard rate. We make the following comments on these results.

(a) In most cases the simulated extra premium is much less than 150%. This suggests that people with a family history of BC/OC are far from being uninsurable.

(b) The dispersion at lower ages (entry age 20) is very small. This is affected by two factors: the EPVs given presence in each state (B_x^{ij} and C_x^{ij} in Equation (6)) and the composition of the underwriting classes (OP_x^{ij} in Equation (6)). At entry age 20, the occupancy probabilities are the same in every simulation, and so is the composition of underwriting classes.

(c) At higher ages different patterns of correlation may affect both the location and the dispersion of the simulated premiums. The means of the simulated premiums are not the same under different patterns of correlation.

(d) All-positive correlation (S1 in Table 4) results in the largest dispersion and the broadest 95% confidence intervals.

(e) For entry age 50 with policy term 10 years, the patterns of correlation do not affect the premium rates for each given genotype but can change the occupancy probabilities OP_x^{ij} at age 50, so differences do appear, unlike the case of known genotype.

6. CONCLUSIONS

Antoniou *et al.* (2003) published relative risks of BC and OC onset in respect of BRCA1 and BRCA2 mutation carriers, with confidence intervals. Their
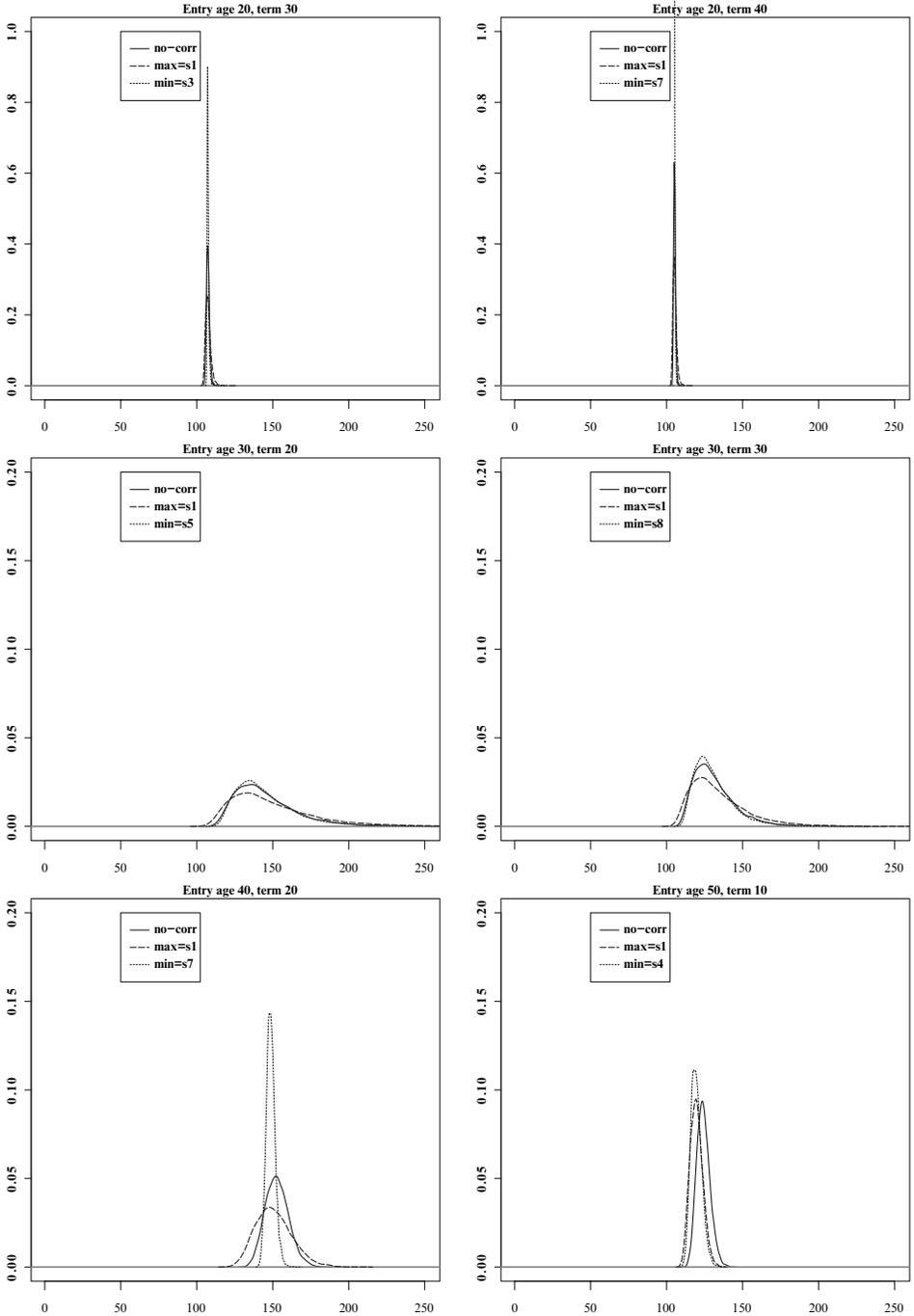
FIGURE 4: Empirical distribution of the simulated FH premium rates, for a person with a family history of BC/OC with various entry ages and policy terms. The x-axis is the level FH premium, as a percentage of the OR premium rate; the y-axis is the density.

parametric model assumed piecewise-constant relative risks over 10-year age groups. It is straightforward to calculate premium rates for mutation carriers, but these must themselves be point estimates, and we would like to know their sampling distributions. We can obtain sampling distributions by bootstrapping, but we know only the sample variances of the log-relative risks, and not the full covariance matrix. Therefore, we must consider the possible effect of the full covariance matrix on the premium rates.

By selecting eight patterns of strong correlation (with absolute values of off-diagonal elements of the correlation matrix set at 0.9) we tested the effect of sampling covariance on the sampling distribution of level premium rates for CI insurance, and found it to be considerable.

The relevance of sampling distributions of premium rates lies in the question of whether genetic information (meaning genetic test results in the U.K., but including family history in some other countries) can ever be used for underwriting. GAIC in the UK will expect to receive evidence that a genetic test is a reliable indicator of increased insurance risk, in deciding whether it may be used. Their methodology is not published, but statistical properties of premium rate estimates may quite possibly be taken into consideration. Therefore methods of estimating, for example, sample variances of premium rate estimates, from the results of published studies, are needed. This is not usually straightforward.

In the case considered here, presence of a mutation or a family history presents convincing statistical evidence to justify charging an extra premium for CI insurance, in all cases. For younger applicants with a family history, however, the increased premiums are so small that they would probably be ignored.

The results shown here may raise the question: what is special about genetics? Why are sampling distributions, or just confidence intervals, not routinely considered in medical underwriting, or quoted where possible in underwriting manuals?

REFERENCES

ANTONIOU, A.C., PHAROAH, P.D.P., MCMULLAN, G., DAY, N.E., STRATTON, M.R., PETO, J., PONDER, B.J. and EASTON, D.F (2002) A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer* **86**, 76-83.

ANTONIOU, A.C., PHAROAH, P.D.P., NAROD, S., RISCH, H.A., EYFJORD, J.E., HOPPER, J.L., LOMAN, N., OLSSON, H., JOHANNSSON, O., BORG, A., PASINI, B., RADICE, P., MANOUKIAN, S., ECCLES, D.M.,

TANG, N., OLAH, E., ANTON-CULVER H., WARNER, E., LUBINSKI, J., GRONWALD, J., GORSKI, B., TULINIUS, H., THORLACIUS, S., EEROLA, H., NEVANLINNA, H., SYRJÄKOSKI, K., KALLIO-NIEMI, O.-P., THOMPSON, D., EVANS, C., PETO, J., LALLOO, F., EVANS, D.G. and EASTON, D.F. (2003) Average risks of breast and ovarian cancer associated with mutations in BRCA1 or BRCA2 detected in case series unselected for family history: A combined analysis of 22 studies. *American Journal of Human Genetics* **72**, 1117-1130.

DAYKIN, C.D., AKERS, D.A., MACDONALD, A.S., MCGLEENAN, T., PAUL, D. and TURVEY, P.J. (2003) Genetics and insurance – some social policy issues (with discussions). *British Actuarial Journal* **9**, 787-874.

EASTON, D.F., BISHOP, D.T., FORD, D., CROCKFORD, G.P. and THE BREAST CANCER LINKAGE CONSORTIUM (1993) Genetic linkage analysis in familial breast and ovarian cancer. *American Journal of Human Genetics* **52**, 678-701.

EASTON, D.F., FORD, D., BISHOP, D.T. and THE BREAST CANCER LINKAGE CONSORTIUM (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. *American Journal of Human Genetics* **56**, 265-271.

GLASSERMAN, P. (2004) *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

GUI, E.H., LU, B., MACDONALD, A.S., WATERS, H.R. and WEKWETE, C.T. (2006) The genetics of breast and ovarian cancer III: A new model of family history. *Scandinavian Actuarial Journal* **2006**, 338-367.

GUTIÉRREZ, M.C. and MACDONALD, A.S. (2003) Adult polycystic kidney disease and critical illness insurance. *North American Actuarial Journal* **7(2)**, 93-115.

LU, L. (2006) Some actuarial and statistical investigations into topics on genetics and insurance. Ph.D. dissertation, Heriot-Watt University, Edinburgh.

LU, L., MACDONALD, A.S. and WATERS, H.R. (2007) A dynamic family history model of hereditary nonpolyposis colorectal cancer. *To appear in Annals of Actuarial Science*.

LU, L., MACDONALD, A.S. and WEKWETE, C.T. (2007) Premium rates based on genetic studies: How reliable are they? *Insurance: Mathematics & Economics* **42**, 319-331.

MACDONALD, A.S., WATERS, H.R. and WEKWETE, C.T. (2003) The genetics of breast and ovarian cancer I: A model of family history. *Scandinavian Actuarial Journal* **2003**, 1-27.

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