A.A.S. **1**, II, 319-343 (2006)

APPLICATION OF A POLYGENIC MODEL OF BREAST AND OVARIAN CANCER TO CRITICAL ILLNESS INSURANCE

By A. S. Macdonald and K. R. McIvor

ABSTRACT

Mutations in the BRCA1 and BRCA2 genes confer very high risk of breast cancer (BC), but only account for about 25% of the observed familial clustering of BC. Antoniou *et al.* (2002) proposed a model which included the BRCA1 and BRCA2 genes, and a polygenic component which acted multiplicatively on the rate of onset of BC. We use this model to find premium rates for critical illness insurance: (a) given knowledge of an applicant's polygenotype; and (b) given knowledge of a family history of BC or ovarian cancer. We find that the polygenic component causes large variation in premium rates even among non-mutation carriers, therefore affecting the whole population. In some cases the polygenic contribution is protective enough to reduce or remove the additional risk of a BRCA1/2 mutation, leading to cases where it will be advantageous to disclose genetic test results which are adverse in absolute terms. Premiums based on family history are lower than those found in an earlier study which attributed all genetic BC risk to the BRCA1/2 genes.

KEYWORDS

Polygenotype; Breast Cancer; Critical Illness Insurance; BRCA1; BRCA2; Ovarian Cancer

CONTACT ADDRESS

Angus Macdonald, Department of Actuarial Mathematics and Statistics, 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. Introduction

1.1 *Breast Cancer, Ovarian Cancer and Insurance*

Breast cancer (BC) is the most common cancer among women in the United Kingdom; one in nine women develop BC in their lifetime. Ovarian cancer (OC) is the fourth most common cancer among women, and the U.K. has the highest incidence of OC in Europe (Cancer Research U.K.). Together they account for a significant proportion of claims under critical illness (CI) insurance policies. It is well known that mutations in either the BRCA1 or BRCA2 genes can increase the risk of BC or OC at early ages very substantially.

The genetic risk associated with family histories of BC or OC has prompted more actuarial research than has any other genetic disorder. The

work has built upon the genetic epidemiology of BC and OC, which is still developing. Early epidemiological studies selected highly affected families; these were the basis for actuarial studies by Subramanian *et al.* (1999), Lemaire *et al.* (2000) and Macdonald *et al.* (2003a, 2003b). Recent advances in the epidemiology include larger sample sizes and less biased selection of subjects or families. A recent actuarial study allowing for these is Gui *et al.* (2006). The aim of all these actuarial studies has been to model how life and CI insurance pricing may be affected: (a) if the insurer knows of the genetic risk; or (b) if the applicant for insurance knows of the genetic risk, but the insurer does not.

In the U.K. the Genetics and Insurance Committee (GAIC) has the task of assessing applications made by the insurance industry to be allowed to use genetic test results in underwriting, provided: (a) the test results were known because of past clinical history; and (b) the sum assured exceeds the limit set in an agreed moratorium (currently £500,000 for life insurance and £300,000 for CI insurance). Because of their significance, tests for BRCA1/2 mutations are very likely to be the subjects of applications to GAIC.

U.K. insurers are still allowed to use family history in underwriting (unlike in some other countries, such as Sweden), so, in view of the high limits set by the moratorium, the vast majority of applications involving a family history of BC or OC will continue to be underwritten on that basis. Although genetic test results have attracted much attention, the implications of a family history are of more practical importance.

The main epidemiological quantity needed for actuarial modelling is the rate of onset, here denoted $\mu_g(x)$. This is the force of onset of the disease (or hazard rate) at age x, for a person with genotype g. If estimates of $\mu_g(x)$ are available, they can be incorporated in a multiple decrement model for CI insurance almost trivially, or, more generally, given any payment function we can compute its expected present value (EPV), denoted EPV(g). However, this assumes the genotype g to be known. If all that is known is the existence of a family history when a woman aged x applied for insurance, the corresponding EPV is:

$$\sum_g \text{P[Genotype is } g \mid \text{Family history exists at age } x] \, \text{EPV}(g) \qquad (1)$$

where the sum is over all possible genotypes g. Thus, the genotype-specific quantities are still needed, even if the focus is on family history. An important point, which will drive our choice of methodology later, is that the conditional genotype probabilities in equation (1) usually depend on the transmission probabilities, namely the probabilities that a child of parents whose genotypes are known will have any given genotype.

Another key feature of the earlier genetic epidemiology of BC and OC is that it was based upon the inheritance of major genes, namely single genes in

which mutations are alone sufficient to cause the disease. BRCA1 and BRCA2 are the most important, but in the Appendix we list other genes which have been associated with BC risk. Current and future epidemiology is likely to change direction radically, and emphasise another class of genes, called polygenes. Polygenic inheritance means that the genetic risk depends on which variants of several different genes are inherited. In fact, diseases associated with mutations in single genes are exceptional, the vast majority of genetic risks in adult life are almost certainly polygenic, and may be influenced by environment and lifestyle too (hence the name 'multifactorial disorder' which is often used to describe them). Even when major genes may cause a disease, it is possible that the majority of familial clustering of the disease may be caused by polygenes. This is very likely to be the case with BC (see Section 2.1). This epidemiological breakthrough will offer a completely new perspective on the insurance issues raised by knowledge of an individual's genetic profile.

Recently Antoniou *et al.* (2002) examined a number of genetic models for BC/OC risk, using data which included both high-risk families and families not selected for known BC risk. The best-fitting model incorporated BRCA1, BRCA2 and a polygene which modified the rates of onset of BC. Since their paper, and other published sources, allow all the $\mu_g(x)$ to be found, it is simple to build an actuarial model for CI insurance assuming genotypes to be known, hence to answer the question: "What is the effect of pricing CI insurance if the polygene is allowed for, as well as the major genes BRCA1 and BRCA2?" Put another way: "How reliable is a genetic test which shows a BRCA1/2 mutation to be present, if the polygene is *not* taken into account?" This bears directly on the criteria that GAIC has published for assessing the use of genetic tests by insurers.

We then wish to study how the polygene affects CI insurance pricing if, as usual, only the existence of a family history is known. Previous studies have used equation (1) directly, because a small number of major genes defines a small number of genotypes g. This is not the case with the polygenic model, in particular the conditional genotype probabilities in equation (1) are intractable. We therefore simulate a large number of nuclear families, and assume that the children of these families make up the pool of potential applicants for insurance. The empirical distribution of genotypes in this simulated sample provides the probabilities in equation (1) directly.

2. The Model of Antoniou *et al.* (2002)

2.1 *Breast Cancer and Polygenes*

The risks of BC and OC onset were linked to mutations in the BRCA1 and BRCA2 genes in the 1990s, triggering a search for other genes implicated

in tumour formation. This search still goes on, because it is estimated that BRCA1, BRCA2 and the other possible high-risk genes found to date (see the Appendix), account for only about 25% of the observed familial clustering (Struewing, 2004; Easton, 2005). Part of the problem is that mutations in BRCA1/2 are quite rare, their frequencies being estimated to be 0.051% and 0.068% respectively (Antoniou *et al.*, 2002).

It is widely believed that the remaining component arises from the combined influence of common variations (alleles) in several genes which each, individually, has only a small effect on the risk of BC. Such a configuration is called 'polygenic', and the genes which contribute to it may collectively be called a 'polygene'. Although it is unlikely that a polygene explains all of the remaining 75% of the familial variation (there are other shared factors within families, such as diet and socio-economic status), it may explain a larger proportion than do any of the major genes.

2.2 *The Hypergeometric Polygenic Model*

The inheritance of major (single) genes, except those carried on the sex chromosomes, is usually assumed to follow Mendel's laws, summarised as follows: everyone carries two copies of every gene and each of their children inherits one of them, selected randomly and independently. Thus, the chance that a child receives either copy carried by a given parent is 1/2. This is quite tractable if we are interested in a small number of major genes, each with a small number of alleles. For example, if we regard BRCA1 and BRCA2 as each having two alleles (mutated and normal) there are only $3 \times 3 = 9$ possible genotypes, whose frequencies can be calculated exactly if the allele frequencies are known.

However a polygenic model may involve a large number of genes, each with several alleles. In principle Mendel's laws may still be applied, but the number of possible genotypes quickly becomes intractable in many practical problems. For example, if six genes contribute to the polygene, and each has two alleles, there are $3^6 = 729$ possible genotypes. It is impractical to specify the effect of each of these genotypes on disease onset without some making some simplifying assumptions — in other words, a model.

Consequently, approximate models of the polygenic contribution to a disease have been proposed. A widely used assumption is that the polygenotype is represented by a numerical value on a continuous scale, and the distribution of these values in the population is Normal. This can be motivated by applying the Central Limit Theorem to a model in which total disease risk is the sum of the disease risks associated with all the alleles contributing to the polygenotype, with suitable independence assumptions. The polygenic disease risk may then be a suitable function of the polygene's numerical value, or the disease may be assumed to occur if the polygene's value exceeds a threshold. While this gives a simple model of a polygene's

effect, it makes it difficult to model inheritance. The question to be answered is: "What is the conditional probability that the child of parents with known polygenotypes will have any given polygenotype?" Passage to the continuous limit simplifies the problem of having many additive contributions to the risk, but at the same time it turns the combinatorics of inheritance from hard to impossible.

As a result, when the inheritance of a polygene must be modelled, approximations may be made in the other direction, from continuous to discrete. The numerical polygenotype is assumed to take values in a discrete distribution with a suitable shape, for which there is a plausible model of transmission from parents to children. (Note that this discretisation does not mean a return to the Mendelian model; it is not now genes which are transmitted from parents to children but just a numerical 'value' representing the polygene.) Before giving an example, we fix terminology, by making the following conventions:

(a) The word 'polygene' will mean the collection of genes which constitute it — actual physical segments of DNA.
(b) Variants of a gene which contributes to a polygene will be called 'polygenic alleles'.
(c) The word 'polygenotype' means a numerical value representing the polygene.

Our example is the hypergeometric model of Lange (1997), derived from Cannings *et al.* (1978). It was used by Antoniou *et al.* (2002) to represent a polygenic component of BC risk, and it will be central in this paper.

Suppose that n genes, inherited independently of each other, contribute to the polygene, and that each has an 'adverse' allele and a 'beneficial' allele, which are equally common. An adverse allele contributes $+1/2$ to the numerical value of the polygenotype, and a beneficial allele $-1/2$. Since a person has two copies of each gene, the polygene is defined by the total number of adverse alleles, the possibilities being $0, 1, \ldots, 2n$. The corresponding numerical values of the polygenotype are $-n, -(n-1), \ldots, (n-1), n$, meant to suggest that 'negative' polygenotypes present below average risk, while 'positive' polygenotypes present above average risk. The mother's, father's and child's polygenotypes are random variables denoted P_m, P_f and P_c respectively. Assuming the parents to be sampled randomly from the population, their polygenotypes are independently binomially distributed with parameter $(2n, 1/2)$, for example:

$$\mathrm{P}[P_m = p_m] = \binom{2n}{p_m + n}\left(\frac{1}{2}\right)^{2n} \qquad (p_m = -n, -(n-1), \ldots, (n-1), n). \quad (2)$$

Thus the 'extreme' polygenotypes are uncommon, and the 'central'

polygenotypes much more common. By the assumed independence, the parents' joint polygenotype is:

$$\mathrm{P}[P_m = p_m, P_f = p_f] = \binom{2n}{p_m + n}\binom{2n}{p_f + n}\left(\frac{1}{2}\right)^{4n}. \tag{3}$$

Polygenes are transmitted from parents to children by independently sampling, without replacement, n polygenic alleles from the mother and n from the father. Conditional probabilities for an offspring's polygenotype are then:

$$\begin{aligned}&\mathrm{P}[P_c = p_c | P_m = p_m, P_f = p_f] \\ &\quad = \sum_{r=\max[0,\,p_c - p_f]}^{\min[p_m+n,\,p_c+n]} \frac{\binom{p_m+n}{r}\binom{n-p_m}{n-r}}{\binom{2n}{n}} \frac{\binom{p_f+n}{p_c+n-r}\binom{n-p_f}{r-p_c}}{\binom{2n}{n}}.\end{aligned} \tag{4}$$

This is the convolution of two independent hypergeometric distributions representing the sum of the father's and the mother's contributions to their child's polygenotype. For further details see Lange (1997).

2.3 *The Model of Antoniou* et al. *(2002)*

Antoniou *et al.* (2002) fitted several alternative models to a set of high-risk families (each with multiple cases of BC or OC) and a set of unselected BC cases. The best-fitting model was a mixed major gene and polygenic model, in which the major genes were BRCA1 and BRCA2. The site of a mutation on BRCA1/2 was not considered; mutations were either present or absent. Previous studies have shown different mutation sites on the BRCA genes to display different risks of onset and aggressiveness after onset, but this aspect of the epidemiology of BC/OC is not yet developed enough to be taken into account.

For convenience, we use the term 'BRCA0 genotype' to indicate a person who carries neither BRCA1 nor BRCA2 mutations, and let 'BRCA1 genotype' and 'BRCA2 genotype' refer to mutation carriers, although strictly there is no such gene as BRCA0.

The authors used the national incidence rates for England and Wales in 1983 to 1987 as baselines, and estimated the relative risks of BC and OC in respect of BRCA1 and BRCA2 mutation carriers, piecewise constant over ten-year age groups between ages 30 and 69. These are shown in Table 1. Since they did not publish the baseline rates, we calculated our own using ONS statistics for England and Wales in 1983 to 1987 and cancer registrations over the same period (ONS, 1999). These are shown in Figure 1, along with

Table 1. The relative risks for BC and OC BRCA1 or BRCA2 mutation carriers estimated by Antoniou *et al.* (2002); the baselines are the onset rates in England and Wales in 1983 to 1987

	Breast cancer		Ovarian cancer	
Age	BRCA1	BRCA2	BRCA1	BRCA2
30-39	23.88	17.52	3.43	3.67
40-49	12.40	10.80	53.32	2.00
50-59	4.91	12.11	20.86	11.85
60-69	2.31	12.53	19.51	8.32

Table 2. Comparison of the incidence rates for breast cancer estimated by Antoniou *et al.* (2002) and Ford *et al.* (1998)

	Antoniou *et al.*		Ford *et al.*	
Age	BRCA1	BRCA2	BRCA1	BRCA2
30-39	0.011222	0.008236	0.01618	0.0118
40-49	0.016621	0.014471	0.04749	0.0210
50-59	0.008255	0.020352	0.03480	0.0318
60-69	0.004843	0.026326	0.02162	0.1180

crude estimates of those used by Antoniou *et al.* (2002), obtained by dividing absolute onset rates by the relative risks. Thus we have onset rates $\mu^{BC}_{BRCAi}(x)$ and $\mu^{OC}_{BRCAi}(x)$, for $i = 0, 1, 2$.

Table 2 compares the BC incidence rates of BRCA1 and BRCA2 mutation carriers from this study with those of the earlier study by Ford *et al.* (1998) (the basis of the actuarial model of Macdonald *et al.* (2003a)). The trends with age are similar, but the rates from Antoniou *et al.* (2002) are much lower, particularly for older BRCA2 mutation carriers. This is as expected, because Ford *et al.* (1998) included only high-risk families (those with at least four cases of BC), whereas Antoniou *et al.* (2002) included a population-based cohort. Both studies focused on early onset of BC, with relatively few cases of onset at ages over 50 to 55, possibly leading to underestimated risk at higher ages.

The polygenotype is modelled as a Normal random variable R with mean zero and variance (the fitted parameter) $\sigma_R^2 = 1.291$. It modifies the BC risk regardless of BRCA genotype as follows:

$$\mu^{BC}_{BRCAi}(x, R) = \mu^{BC}_{BRCAi}(x)e^R. \tag{5}$$

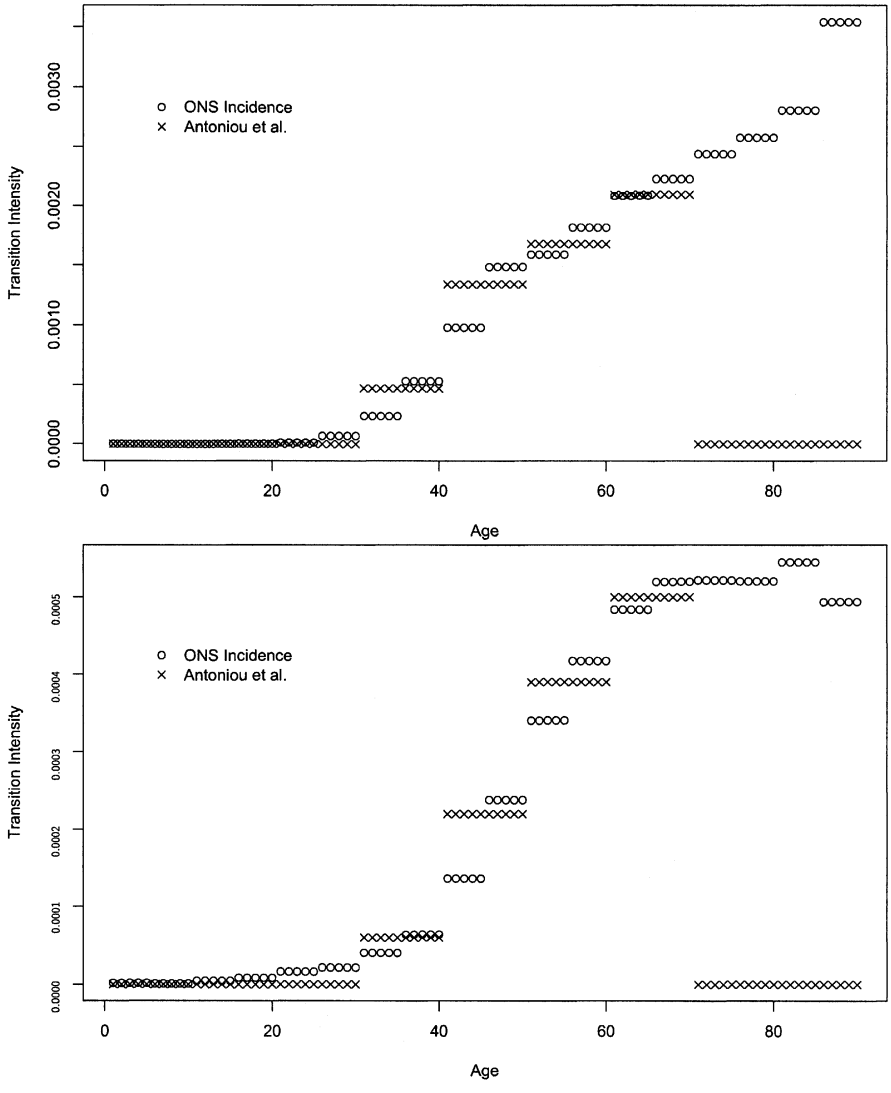

Figure 1. Baseline incidence rates for BC (top) and OC (bottom) from ONS figures for England and Wales (1983 to 1987) and figures from Antoniou *et al.* (2002)

The Normal polygenic model was discretised to calculate likelihoods. Antoniou *et al.* (2002) used the hypergeometric model (Section 2.2) with $n = 3$, thus seven polygenotypes P, with values $-3, -2, -1, 0, 1, 2, 3$, binomially distributed as in equation (2). Values of R were approximated in terms of values of P (equating second moments) as follows:

$$R \approx \frac{P}{\sqrt{n/2}\sigma_R}. \tag{6}$$

The polygenotype did not affect the incidence of OC, or of any other disorder.

As we will need to model the transmission of polygenotypes from parents to children, we will use the same model.

3. A Model for Critical Illness Insurance

3.1 *The Model*

Figure 2 shows a continuous-time Markov model of a CI insurance contract. The transition intensities from 'healthy' to 'other critical illness' and 'dead' are taken from Gutiérrez & Macdonald (2003).

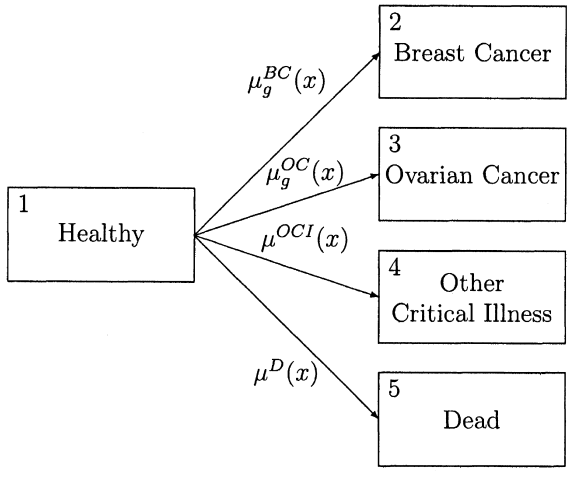

Figure 2. A model of the life history of a critical illness insurance policyholder, beginning in the healthy state; transition to a non-healthy state d at age x is governed by an intensity $\mu^d(x)$ depending on age x or, in the case of BC and OC, $\mu_g^d(x)$ depending on genotype g as well

3.2 *Premiums Based on Known Genotypes*

Table 3 shows the level net rates of premium, payable continuously, for CI insurance cover at several entry ages and policy terms. The premium rates are expressed as a percentage of those for a woman who carries no BRCA1/2 mutation (genotype BRCA0) and who has the 'neutral' polygenotype $P = 0$, which we take to be the 'standard' premium. The force of interest is 0.05 per annum. Expected present values (EPVs) were found numerically by solving Thiele's equations (Hoem, 1988) using a Runge-Kutta algorithm with step size 0.0005 years.

In CI insurance, premiums in excess of 300% to 350% of the standard premium usually result in cover being declined. Many of the ratings for known BRCA1 and BRCA2 mutation carriers are above this level. Previous studies using quite recent epidemiology, but the major genes only, have reported that both BRCA1 and BRCA2 mutation carriers are likely to be declined for any combination of entry age and term (Gui *et al.*, 2006). Our results with the polygene $P = 0$ mostly agree with this.

The variation by polygenotype is the most striking feature of these results, and, since it affects the whole population, not just the carriers of rare mutations, it presents for the first time a widespread major variation of a genetic risk factor:

(a) The polygene alone (genotype BRCA0) leads to premiums for the highest risk ($P = +3$) that are up to 3.4 times those for the lowest risk ($P = -3$). Variation of this order caused by a major gene would probably be worthy of an actuarial study in its own right.
(b) In some instances a BRCA1/2 mutation carrier with a protective polygenotype may be eligible for a lower premium than non-mutation carriers with a risky polygenotype.
(c) We see that BRCA1/2 mutation carriers can be offered CI insurance at most entry ages and policy terms if they have a strongly protective polygenotype. Thus, there is potential for genetic testing to make insurance more accessible under a lenient moratorium (one in which genetic test results may be disclosed if it is to the applicant's advantage).

On the other hand, premiums are even higher than previously reported for women with a detrimental combination of genotypes. The premium rate in the worst case (polygenotype +3 and major genotype BRCA1) is up to 38 times the 'standard' rate and up to 26 times the premium rate for a BRCA1 mutation carrier with polygenotype −3. For BRCA2 mutation carriers the corresponding multiples are about 28 and 22 times.

3.3 *A Comment on Genetic Tests for Polygenotypes*

References to 'known' polygenotypes should not lead readers to suppose they might soon be detected by DNA-based genetic tests. Our model of a

Table 3. Level net premium for women free of BRCA1/2 mutations, depending on polygenotype, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with the mean polygene $P = 0$

Major genotype	Polygenotype	Age 20				Age 30			Age 40		Age 50
		10 years %	20 years %	30 years %	40 years %	10 years %	20 years %	30 years %	10 years %	20 years %	10 years %
BRCA0	−3	94.5	87.1	83.7	85.2	83.0	81.4	83.9	80.3	83.8	86.5
	−2	95.3	88.8	85.9	87.2	85.4	84.0	86.1	83.0	86.0	88.3
	−1	96.8	92.5	90.5	91.5	90.2	89.2	90.7	88.5	90.6	92.2
	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	+1	106.4	115.4	119.2	117.3	120.2	122.0	119.0	123.4	119.2	116.1
	+2	119.8	147.0	158.0	151.9	161.6	166.7	157.3	171.4	158.5	149.1
	+3	146.9	211.1	235.3	219.2	246.4	256.8	233.2	269.5	238.2	216.7
BRCA1	−3	94.5	124.5	202.1	181.4	140.4	225.5	194.8	283.7	217.5	160.5
	−2	95.3	160.9	234.4	203.8	196.2	265.7	221.4	317.0	238.1	169.6
	−1	96.8	234.9	299.1	248.6	310.2	347.3	275.8	385.0	280.4	188.3
	0	100.0	383.3	425.6	336.1	542.5	511.7	386.2	523.8	367.8	226.7
	+1	106.4	673.8	660.9	497.9	1,012.5	837.3	607.4	805.1	549.0	305.2
	+2	119.8	1,215.7	1,057.5	767.3	1,949.6	1,460.1	1,037.6	1,368.7	927.1	465.6
	+3	146.9	2,130.8	1,615.8	1,138.8	3,765.0	2,585.9	1,827.3	2,474.5	1,714.2	792.1
BRCA2	−3	94.5	115.7	111.9	122.2	126.9	115.6	125.9	108.9	126.2	142.8
	−2	95.3	142.6	139.3	146.2	167.8	148.9	153.5	138.2	151.2	165.3
	−1	96.8	197.3	194.5	193.7	251.6	216.6	209.0	198.2	202.0	211.5
	0	100.0	307.7	303.1	284.3	422.7	353.1	318.7	320.6	304.3	306.0
	+1	106.4	526.6	508.0	445.1	770.1	623.3	527.9	569.1	507.1	498.8
	+2	119.8	945.5	862.7	696.3	1,467.8	1,139.4	903.3	1,068.3	896.7	890.5
	+3	146.9	1,690.3	1,385.9	1,023.5	2,839.6	2,066.6	1,532.3	2,051.4	1,616.0	1,680.6

polygenotype is a numerical value, whereas a real polygenotype is a combination of (possibly very many) alleles. In order to test for a polygene and relate the result to a risk estimate, all the complications that drive geneticists to use the simplified model will have to be overcome. Moreover, it seems unlikely that genetic risks will be capable of being understood in isolation, but only in combination with other major risk factors.

4. Modelling Family History in Insurance Underwriting

4.1 *Modelling Family History*

The problem is to find EPVs given a family history at age x, as in equation (1). Assuming the genotype-specific onset rates to be known, this reduces to estimating the conditional probabilities:

$$\text{P[Genotype is } g \mid \text{Family history exists at age } x]. \tag{7}$$

First, we must define what is meant by 'family history'. That done, the calculation must be anchored by the assumption that some ancestors of the applicant have genotypes which are randomly and independently sampled from the distribution of genotypes in the population. We will assume this to be true of the applicant's parents; thus their genotype probabilities are known. Together with the transmission probabilities which govern the inheritance of genes, this fixes the genotype probabilities of the applicant and all her siblings. For every possible joint genotype of the entire family, we know the probabilities of critical illnesses, including BC and OC, striking before any given age, hence the probability of a family history arising. At this point, the computation of the probability (7) has become, in principle, just an application of Bayes' Theorem. However, the summation is not over the applicant's possible genotypes as in equation (1), but over all possible joint genotypes of the whole family.

The procedure outlined above was followed by Macdonald *et al*. (2003a) for several definitions of family history. They also considered the more realistic possibility that the insurer may not have any information about the unaffected relatives of the applicant. Their approach could not be extended to a model of the insurance market, necessary to study the potential costs of adverse selection, because it did not model the development of a family history over time as a factor which might influence the decision to buy insurance or to take a genetic test. That step was taken by Gui *et al*. (2006), who pointed out that, if the definition of 'family history' is such that at any given time it is either certainly present or certainly absent, the time at which it appears can be modelled as an event time in the usual framework of survival models, and the procedure outlined above can be modified to give an age-dependent 'rate of onset' of a family history. However, this approach

still depended on applying Mendelian transmission probabilities to just two major genes.

The polygenotype introduces a non-Mendelian model of transmission, which is not a real problem, and greatly increases the number of genotypes, which is. Thus, we have chosen to estimate the probabilities (7) by simulation.

4.2 *Definition of Family History*

Our definition of a family history is based on a typical underwriting threshold, namely two first-degree relatives (FDRs, meaning parents and siblings) suffering onset of BC or OC before age 50. Under many underwriting standards this condition would lead to an extra premium being charged (Macdonald *et al.*, 2003b). Note that this is quite different from clinical practice, in which a family history may be defined by a much more complex pedigree, including second-degree and other relatives. To a clinician, also, a family history is defined by the circumstances of each patient. Thus we rely on the much simpler notion used by insurers.

4.3 *The Simulation Model*

The approach is as follows:

(a) A family starts with two parents, whose major genotypes and polygenotypes are independently sampled from their respective distributions in the population, except that we disregard the probability that either parent has more than one mutation. This is consistent with the treatment of BRCA1 and BRCA2 in Antoniou *et al.* (2001). It is widely assumed by epidemiologists that a foetus with two mutations of the same BRCA gene will not be viable and will miscarry. We use the BRCA1 and BRCA2 mutation frequencies from the polygenic model in Antoniou *et al.* (2002), 0.051% and 0.068% respectively.

(b) The number of daughters which the parents have is randomly sampled from a suitable distribution. We use that of Macdonald *et al.* (2003a), which is given in Table 4. Hence the family size may vary from three to nine members, and the father is the only male. For simplicity, we assume that the mother has her children when she is age 30 and that all daughters are the same age.

Table 4. Distribution of the number of daughters born in a family

No. of daughters	Probability	No. of daughters	Probability
1	0.54759802	5	0.00285702
2	0.33055298	6	0.00035658
3	0.09749316	7	0.00002634
4	0.02111590		

Source: Macdonald *et al.* (2003a).

(c) Each daughter, independently of the others, inherits the major genes at random according to Mendel's laws, and the polygenotype at random according to equation (4). We discard any family in which a daughter inherits any two major gene mutations, for the reasons given in (a).
(d) The life histories of the mother and daughters, in respect of the model in Figure 2, are simulated using a competing risks approach. We ignore male BC, and we assume that the mother is healthy at age 30.

After simulating a large number of such families, we can observe, at every age $x > 0$, the distribution of the genotypes of daughters in families in which a family history has appeared. We will describe this in Section 4.6.

4.4 *Simulating Competing Risks*

There are four decrements in the model in Figure 2. Define T^{id} to be the random time at which the ith person in the simulated sample suffers decrement d, as if it acted alone. In the simulation, the ith person's genotype is known, say it is g. Then T^{id} has distribution function, denoted $F_g^d(t)$, given by:

$$F_g^d(t) = 1 - \exp\left(-\int_0^t \mu_g^d(x+s)ds\right)$$

for $t < \infty$, possibly with a probability mass at $t = \infty$. This, and its inverse, can be computed and tabulated. The random variable $F_g^d(T^{id})$ is uniformly distributed on $[0, 1]$, so we simulate a uniform $[0, 1]$ random variable, denoted a^{id}, and solve numerically the equation $F_g^d(t^{id}) = a^{id}$ to obtain our simulated value t^{id}. The ith person's life history is then represented by the pair (t^i, d^i), where $t^i = \min[t^{i1}, t^{i2}, t^{i3}, t^{i4}]$ and d^i is that decrement for which $t^{ij} = t^i$.

Note that each decrement in the model censors the others, so it is not possible for a woman who survives a heart attack (for example) to develop BC/OC subsequently. The effect is minimal at those ages where onset would contribute to a family history; by age 50 only about 6% of women have developed one of the other CIs.

4.5 *Sampling Insurance Applicants from Simulated Families*

We simulated 10,000,000 families as described above, containing in total 16,022,024 daughters. At any age x, those daughters still healthy constitute the pool of potential applicants for insurance. We assume that the insurer, in effect, samples randomly from this pool, knowing only whether each applicant has a family history or not. As well as using the maximum possible amount of information in the simulated families, this sampling scheme accounts correctly for the fact that there are more potential applicants than

there are family histories; in larger families the appearance of a family history will affect more than one healthy daughter.

4.6 *Applicant's Genotype Distribution*

We can now estimate by direct enumeration the distribution of the applicant's genotype, conditional on the observed family history. All applicants are healthy, but some have a family history and others do not. This is all that the insurer knows. We, however, also know into which of the following categories each applicant falls:

(a) applicant is in a BRCA0 family (no mutations) and has BRCA0 genotype;
(b) applicant is in a BRCA1 family and has BRCA0 genotype;
(c) applicant is in a BRCA1 family and has BRCA1 genotype;
(d) applicant is in a BRCA2 family and has BRCA0 genotype; or
(e) applicant is in a BRCA2 family and has BRCA2 genotype.

Table 5 shows the numbers of daughters who have no family history at selected ages from zero to 60 years, grouped into the five categories above and the state occupied in the CI model (Figure 2). Table 6 shows the corresponding distribution of daughters who do have a family history. In both tables the potential insurance applicants are those in the healthy state.

We further subdivide the numbers in Tables 5 and 6 by polygenotype. The results are too extensive to tabulate, so, for illustration, Figures 3 and 4 show histograms of the polygenotype distribution among healthy daughters with a family history, for the five major gene categories above, and ages 30 and 40 (Figure 3) and 50 and 60 (Figure 4). Note that no mutation carriers have a family history at age 30. This is because mutation carriers are rare, and before age 30 they share the population onset rates of BC and OC.

For brevity, we omit the polygenotype distributions of daughters with no family history. They are slightly more inclined to less risky values, because carriers of more dangerous polygenotypes are more likely to have FDRs with risky polygenotypes, hence to have a higher risk of developing a family history. This is most pronounced in BRCA2 mutation carriers, because the deleterious effects of BRCA2 mutations are relatively late acting.

These empirical distributions (at all ages x, not just the selected ages illustrated) provide the conditional probabilities which we need (equation (7)) to calculate premiums for a daughter with a family history.

4.7 *Premiums for an Applicant with a Family History*

Sample level premiums for a daughter with a family history, applying for level CI insurance, are shown in the first two lines of Table 7. They are expressed as percentages of the relevant premium for a woman with major gene BRCA0 and polygenotype $P = 0$.

Table 7 shows the effect of allowing for or ignoring the polygene. The full model, labelled 'P + MG', uses both polygene and major gene probabilities in

Table 5. Numbers of daughters with no family history and given major genotype, in each state in the CI model (see Figure 2), at selected ages

Genotype			Daughters' ages						
Family	Applicant	State	0	10	20	30	40	50	60
BRCA0	BRCA0	Healthy	15,946,208	15,891,435	15,831,959	15,693,802	15,299,382	14,294,637	12,524,593
BRCA1	BRCA0	Healthy	16,044	15,983	15,916	15,790	15,273	13,745	12,096
BRCA1	BRCA1	Healthy	16,181	16,025	15,963	15,819	13,316	9,184	7,009
BRCA2	BRCA0	Healthy	21,845	21,766	21,680	21,471	20,807	19,063	16,742
BRCA2	BRCA2	Healthy	21,746	21,583	21,498	21,300	18,723	14,793	10,300
BRCA0	BRCA0	BC	0	476	830	10,196	111,201	360,290	691,286
BRCA1	BRCA0	BC	0	16	17	25	104	249	518
BRCA1	BRCA1	BC	0	76	76	88	1,710	2,682	3,403
BRCA2	BRCA0	BC	0	13	13	19	120	347	790
BRCA2	BRCA2	BC	0	75	75	91	1,840	3,389	6,055
BRCA0	BRCA0	OC	0	213	1,275	4,268	13,633	42,105	94,767
BRCA1	BRCA0	OC	0	2	3	6	8	25	74
BRCA1	BRCA1	OC	0	29	29	31	59	848	1,494
BRCA2	BRCA0	OC	0	1	1	4	13	43	100
BRCA2	BRCA2	OC	0	8	12	18	53	87	636
BRCA0	BRCA0	Other CI	0	10,820	41,246	125,774	352,670	935,951	2,122,256
BRCA1	BRCA0	Other CI	0	12	46	120	331	870	2,009
BRCA1	BRCA1	Other CI	0	10	46	130	375	770	1,450
BRCA2	BRCA0	Other CI	0	12	60	204	542	1,298	2,858
BRCA2	BRCA2	Other CI	0	17	61	191	484	1,107	2,196
BRCA0	BRCA0	Dead	0	43,264	70,898	112,102	163,395	255,082	455,163
BRCA1	BRCA0	Dead	0	31	62	103	157	260	452
BRCA1	BRCA1	Dead	0	41	67	113	167	227	355
BRCA2	BRCA0	Dead	0	53	91	147	210	314	575
BRCA2	BRCA2	Dead	0	63	100	146	211	284	473

Table 6. Numbers of daughters with a family history and given major genotype, in each state in the CI model (see Figure 2), at selected ages

Genotype			Daughters' ages						
Family	Applicant	State	0	10	20	30	40	50	60
BRCA0	BRCA0	Healthy	0	0	0	35	2,771	23,878	18,437
BRCA1	BRCA0	Healthy	0	0	0	0	137	656	551
BRCA1	BRCA1	Healthy	0	0	0	0	128	432	302
BRCA2	BRCA0	Healthy	0	0	0	0	119	570	481
BRCA2	BRCA2	Healthy	0	0	0	0	106	456	225
BRCA0	BRCA0	BC	0	0	0	25	2,801	29,734	32,886
BRCA1	BRCA0	BC	0	0	0	0	22	154	181
BRCA1	BRCA1	BC	0	0	0	0	410	1,585	1,654
BRCA2	BRCA0	BC	0	0	0	0	20	148	180
BRCA2	BRCA2	BC	0	0	0	0	313	1,549	1,730
BRCA0	BRCA0	OC	0	0	0	6	257	2,677	3,008
BRCA1	BRCA0	OC	0	0	0	0	3	20	24
BRCA1	BRCA1	OC	0	0	0	0	11	409	443
BRCA2	BRCA0	OC	0	0	0	0	4	9	10
BRCA2	BRCA2	OC	0	0	0	0	14	40	63
BRCA0	BRCA0	Other CI	0	0	0	0	68	1,406	3,054
BRCA1	BRCA0	Other CI	0	0	0	0	7	49	112
BRCA1	BRCA1	Other CI	0	0	0	0	3	36	59
BRCA2	BRCA0	Other CI	0	0	0	0	6	42	89
BRCA2	BRCA2	Other CI	0	0	0	0	1	31	54
BRCA0	BRCA0	Dead	0	0	0	0	30	448	758
BRCA1	BRCA0	Dead	0	0	0	0	2	16	27
BRCA1	BRCA1	Dead	0	0	0	0	2	8	12
BRCA2	BRCA0	Dead	0	0	0	0	4	11	20
BRCA2	BRCA2	Dead	0	0	0	0	1	10	14

weighting EPVs. The major-gene-only model, labelled 'MG', uses only the major gene probabilities, assuming that everyone has polygenotype $P = 0$. The latter are very much lower, but this has to be interpreted with care:

(a) The major-gene-only model is *not* comparable with previous actuarial studies of CI insurance which were based on the major genes only. Here it just isolates the contribution of the major genes to the familial risk, in the full model. The earlier studies were based on genetic models in which 100% of the familial risk was attributed to the major genes.
(b) What these figures *do* show, in comparison with the earlier studies, is that the larger proportion of the genetic risk of BC/OC lies with the polygene, not with the major genes. This is a very significant conclusion, because genetic testing for the major genotypes is common, but there is no immediate prospect of defining and testing for polygenotype.

Under the major-gene-only model, policies taken out at age 20 have almost

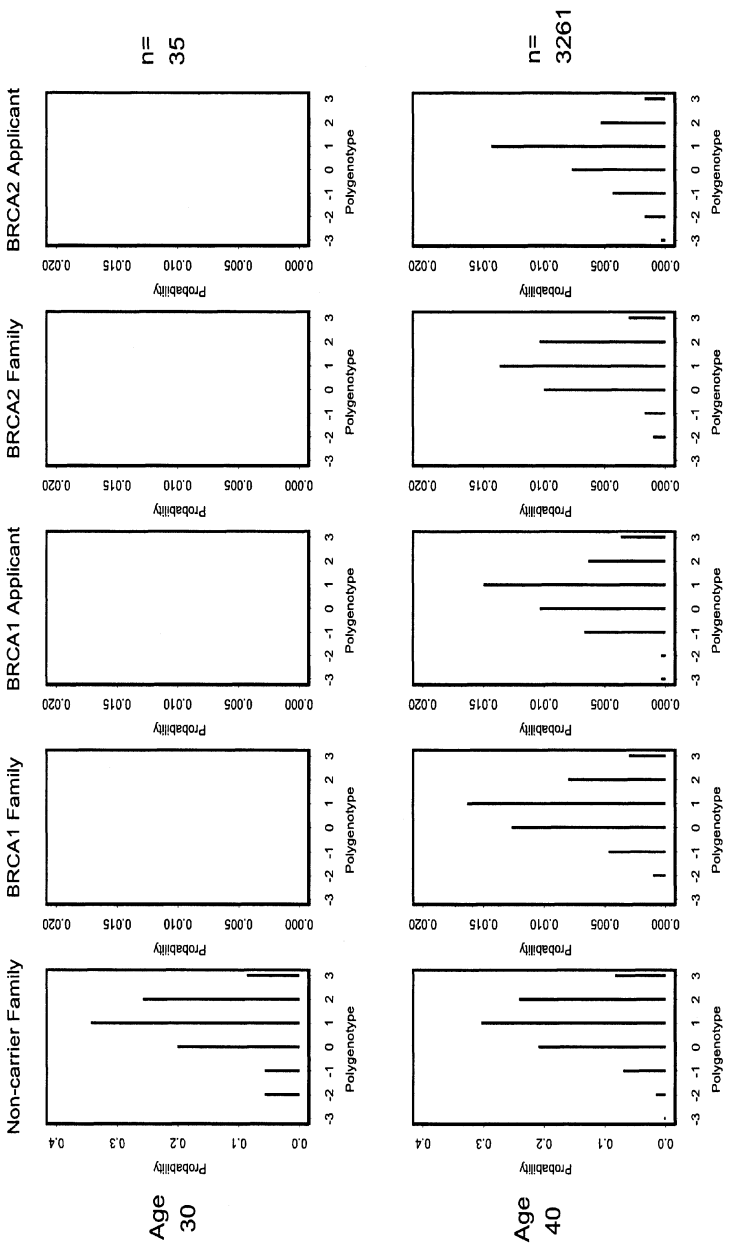

Figure 3. Distribution of polygenotypes by major genotype among healthy daughters aged 30 and 40, with a family history; based on 10,000,000 simulated families; the total number of individuals is shown on the right — note the different vertical scale for non-carrier families

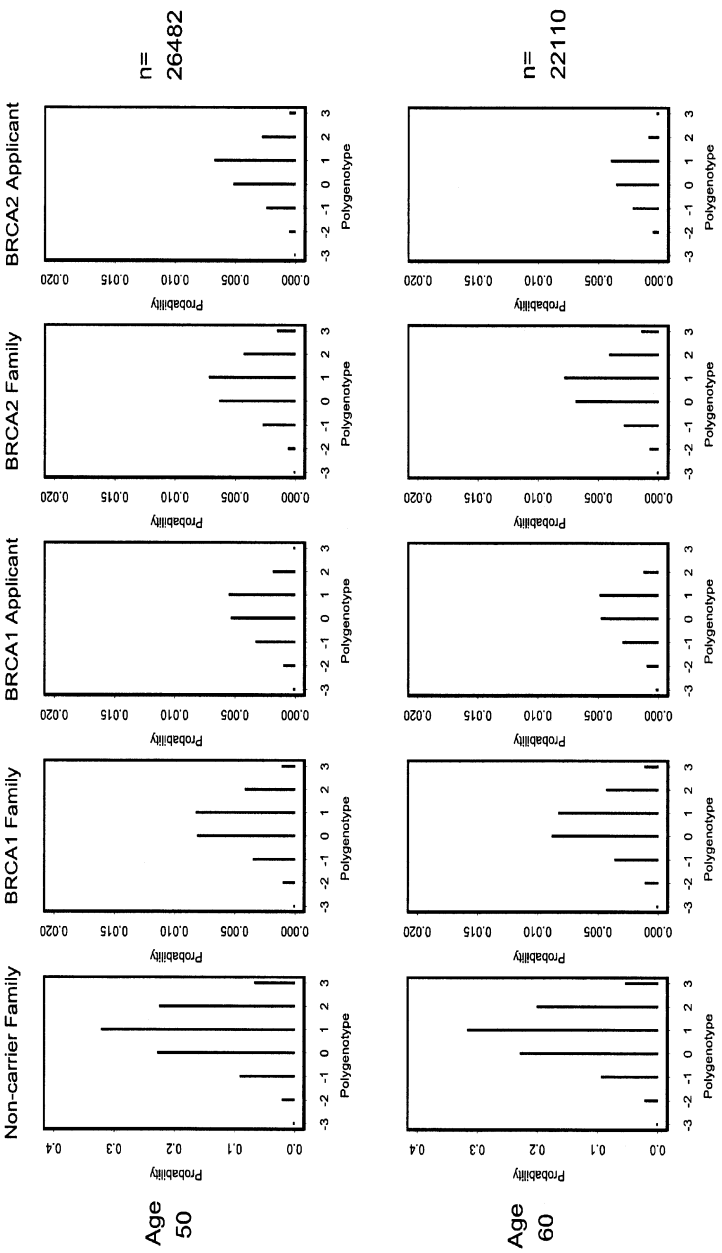

Figure 4. Distribution of polygenotypes by major genotype among healthy daughters aged 50 and 60, with a family history; based on 10,000,000 simulated families; the total number of individuals is shown on the right — note the different vertical scale for non-carrier families

Table 7. Level net premium for females with a family history of BC or OC, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with polygenotype $P = 0$; the P + MG model uses both major gene and polygene probabilities in the weighted average EPVs, while the MG model uses only the major gene probabilities

Definition of family history	Genetic model	Age 30			Age 40		Age 50
		10 years %	20 years %	30 years %	10 years %	20 years %	10 years %
2 affected FDRs	P + MG	134	136	131	175	154	131
	MG	100	100	100	122	115	105
3 affected FDRs	P + MG	100	100	100	293	229	156
	MG	100	100	100	181	157	119
4 affected FDRs	P + MG	100	100	100	381	285	191
	MG	100	100	100	210	174	134

no additional risk because the probability of having developed a family history by age 20 is almost zero (which is consistent with Figure 3).

The premium increases shown under the full polygenic model (P + MG) range from 30% to 75%. The insurer probably would charge an extra premium given these results, but they are quite modest. Clearly, this is a consequence of the definition of family history. We would expect stricter definitions to pinpoint the presence of major genes more accurately, though in a much reduced number of families. Table 7 also shows the increased premiums if a family history is defined as at least three or as at least four first-degree relatives with BC or OC before age 50. As expected they are much higher, in some cases approaching the limit of insurability. However, such family histories are so rare before age 30, even among 10,000,000 simulated families, that the additional premiums were zero for policies taken out at that age.

Macdonald *et al.* (2003b) and Gui *et al.* (2006) gave premium ratings for CI insurance in the presence of a family history of BC or OC. Both used major-gene-only models of BRCA1 and BRCA2, the former based on the study of highly selected families by Ford *et al.* (1988), the latter on a more recent study by Antoniou *et al.* (2003). Moreover, Gui *et al.* used the same definition of family history as we have, namely two FDRs affected before age 50. Table 8 compares our premium rates with theirs, all as percentages of the standard premium. Although Gui *et al.* (2006) was based on a relatively unselected population, they assumed that the onset rates of BC and OC among BRCA1/2 mutation carriers were either 100% or 50% of the rates estimated, as a rough allowance for any remaining ascertainment bias; both are shown in the table.

Our full model (P + MG) yields lower premiums, compared with Gui *et al.* if onset rates were 100% of those estimated. This is as expected for the

Table 8. Level net premium for females with a family history of BC or OC, as a percentage of the standard premium; the polygenic model is compared with the major-gene-only model of Gui *et al.* (2006); the latter assuming that onset rates of BC and OC among BRCA1/2 mutation carriers were either 100% or 50% of those estimated, as a rough allowance for ascertainment bias

Definition of family history	Genetic model	Age 30			Age 40		Age 50
		10 years %	20 years %	30 years %	10 years %	20 years %	10 years %
2 affected FDRs	P + MG	134	136	131	175	154	131
	MG	100	100	100	122	115	105
Gui *et al.* (2006)	100%	330	251	204	208	174	142
	50%	217	179	156	154	139	120

following reason. If we attribute all the inherited BC or OC cases to BRCA1/2 mutations, we will estimate a higher frequency of such mutations in the population, and increase the probability of finding a mutation carrier in a family with a history of BC or OC. Finding a single mutation carrier puts all the individuals, except confirmed non-carriers, at high risk; hence premiums based on a family history are high. By including the polygene we reduce the estimated frequency of BRCA1/2 mutations. Moreover, although a family history may indicate the presence of a dangerous polygene in a parent, their children will inherit an equally dangerous polygene with quite small probability. Polygenotypes, and the physical manifestations which they cause, display classical regression to the mean; indeed, such patterns of inheritance are the very origin of the statistical term 'regression'.

However, it is interesting that our model sometimes yields higher premiums than Gui *et al.* if their onset rates were only 50% of those estimated. Since there is no objective measure of how much their onset rates or ours may have been affected by ascertainment bias, we tentatively conclude that our results show that polygenic inheritance dilutes the strong signals which would be given by family histories if major genes only are responsible.

5. Conclusions

5.1 *The Implications of Polygenes*

This is the first actuarial study to incorporate a fitted model of a polygenic disorder. The following conclusions might be relevant to GAIC when reviewing applications to use genetic test results for BC, or other polygenic diseases, in insurance underwriting.

Very substantial variation in premiums is attributable to the polygenic component of BC and OC risk, as opposed to the much-studied BRCA1 and BRCA2 major genes. Most significantly, some BRCA1/2 mutation carriers could be offered the standard premium rate after a genetic test which accounts for polygenotype. In the context of a lenient moratorium such as that in the U.K. (that is, a moratorium which allows insurers to use a genetic test result if it is to the applicant's advantage), this raises the possibility that a counteracting polygene configuration could be used to void a known BRCA1/2 mutation. At this stage, this is a brave extrapolation from a theoretical polygenic model, but enough genetic variation is unaccounted for by BRCA1 and BRCA2 to make such a conclusion plausible, if and when polygenes become a therapeutic target for BC.

The polygenotype variation in the population (particularly, owing to its size, the subpopulation carrying no BRCA1/2 mutation) could raise questions which have so far largely been avoided because of the rarity of single-gene late-onset disorders. There appears to be enough variation in the risk attributed to the polygenotype that a test for an individual's polygenotype would raise new issues of adverse selection in the insurance market. This will be the subject of future research.

However, our results are consistent with those of Macdonald *et al.* (2003b) and Gui *et al.* (2006), in showing that knowing of a BRCA1/2 mutation only (averaging over polygenotypes) presents a risk high enough to justify increased premiums, beyond the limits of any moratorium which may be in force. Although more recent epidemiology of BRCA1 and BRCA2 have suggested lower penetrance than originally estimated, the fact remains that BRCA1/2 mutation carriers are exposed to a much higher risk of BC and OC.

Because much of the genetic variation in BC can be explained by polygenes which affect the entire population (rather than just mutation carrier families), and the mode of transmission is not Mendelian, a woman with a family history need not have a genotype close to that of her sister. For example, parents with polygenotypes $(P_m, P_f) = (0, 0)$ can produce a child with polygenotype $P_c = +3$ with the same probability as they can produce a child with polygenotype $P_c = -3$. One sister at high risk of BC does not make it certain that any of her sisters will be also. Thus, when we use different models of inherited BC risk we find different premium ratings for a family history. We have also found a large difference in premium ratings if the definition of family history is tightened. Possibly ≥ 3 affected members rather than ≥ 2 affected members is the reasonable threshold of serious risk beyond which insurance may not be attainable.

5.2 *Polygenic Models in Other Diseases*

The genetics and insurance debate has mainly focused, for good reasons, on monogenic disorders, a prime example being Huntington's disease. Now

genetic technologies are advancing rapidly, and we must broaden the focus to include polygenic disorders. This is a more significant undertaking than might first be thought, since every conceivable disorder can be considered to be, to some degree, polygenic. This includes the common disorders like heart disease, cancers and autoimmune diseases. Many common diseases show familial inheritance, but no single genes have been found to account for this.

Interactions between genes and environmental factors make it difficult to identify polymorphisms which influence common diseases. However, large-scale studies, such as U.K. Biobank, are now setting out to map the links between genes and environment. Medical benefits are not expected to appear for at least ten years. When results do begin to come through, however, it is likely that we will find common low-risk genes (polygenes) which are risk factors for a variety of common disorders. It is a prudent pre-emptive step to try to understand the effect which identified polygenes may have on insurance markets. This paper has made some progress towards that.

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. KM is funded by the Engineering and Physical Sciences Research Council.

References

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Ponder, B.J. & Easton, D.F. (2001). Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology*, **21**, 1-18.

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J. & 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.P.D., 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., Kallioniemi, O.-P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D.G. & 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.

Cannings, C., Thompson, E.A. & Skolnick, M.H. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability*, **10**(1), 26-61.

Easton, D.F. (1999). How many more breast cancer predisposition genes are there? *Breast Cancer Research*, **1**, 14-17.

EASTON, D.F. (2005). Finding new breast cancer genes. Presentation at the University of Sheffield.

FORD, D., EASTON, D.F., STRATTON, M., NAROD, S., GOLDGAR, D., DEVILEE, P., BISHOP, D.T., WEBER, B., LENOIR, G., CHANG-CLAUDE, J., SOBOL, H., TEARE, M.D., STRUEWING, J., ARASON, A., SCHERNECK, S., PETO, J., REBBECK, T.R., TONIN, P., NEUHAUSEN, S., BARKARDOTTIR, R., EYFJORD, J., LYNCH, H., PONDER, B.A.J., GAYTHER, S.A., BIRCH, J.M., LINDBLOM, A., STOPPA-LYONNET, D., BIGNON, Y., BORG, A., HAMANN, U., HAITES, N., SCOTT, R.J., MAUGARD, C.M., VASEN, H., SEITZ, S., CANNON-ALBRIGHT, L.A., SCHOFIELD, A., ZELADA-HEDMAN, M. AND THE BREAST CANCER LINKAGE CONSORTIUM (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *American Journal of Human Genetics*, **62**, 676-689.

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

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

HOEM, J.M. (1988). The versatility of the Markov chain as a tool in the mathematics of life insurance. *Transactions of the 23rd International Congress of Actuaries, Helsinki*, **S**, 171-202.

LANGE, K. (1997). An approximate model of polygenic inheritance. *Genetics*, **147**, 1423-1430.

LEMAIRE, J., SUBRAMANIAN, K., ARMSTRONG, K. & ASCH, D.A. (2000). Pricing term insurance in the presence of a family history of breast cancer. *North American Actuarial Journal*, **4**, 75-87.

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

MACDONALD, A.S., WATERS, H.R. & WEKWETE, C.T. (2003b). The genetics of breast and ovarian cancer II: A model of critical illness insurance. *Scandinavian Actuarial Journal*, 28-50.

ONS (1999). Cancer 1971-1997. CD-ROM, Office for National Statistics, London.

REBBECK, T.R. (1999). Inherited genetic predisposition in breast cancer. A population-based perspective. *Cancer*, **86**, 2493-2501.

STRUEWING, J.P. (2004). Genomic approaches to identifying breast cancer susceptibility factors. *Breast Disease*, **19**, 3-9.

SUBRAMANIAN, K., LEMAIRE, J., HERSHEY, J.C., PAULY, M.V., ARMSTRONG, K. & ASCH, D.A. (1999). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: the case of life insurance. *The Journal of Risk and Insurance*, **66**, 531-550.

APPENDIX

GENES CONFERRING BC RISK

The genes listed alongside BRCA1 and BRCA2 in Table 9 are candidate polygenes for BC susceptibility. A polymorphism is defined as an allele with a population frequency of at least 1% (less common alleles are more commonly referred to as 'mutations'). Polymorphisms are extremely common in the human genome (200,000-400,000; Easton, 1999) and therefore offer a vast search region for cancer susceptibility polygenes. In 2005 to 2006 there has been an explosion in published research related to polymorphisms associated with BC (and OC). A quick search of a medical research database (Entrez PubMed) reveals 58 papers published between 1 January 2006 and 11 May 2006.

Table 9. List of genes which may confer additional BC risk (Rebbeck, 1999; Easton, 1999); the allele frequencies are for possible risk-conferring polymorphisms estimated from healthy Caucasian control populations and the numbers of distinct mutations are taken from the Human Gene Mutation Database

Gene	Allele frequency	No. of mutations	BC risk
BRCA1	0.051%	741	High
BRCA2	0.068%	500	High
TP53	39%	139	High
PTEN	<0.01%	170	High
MSH2		337	High
ATM	1%	421	Moderate
CYP1A1	3-11%	2	Moderate
CYP2D6	9%	30	Low
CYP2E1	7-9%	2	Low
CHEK2	1.1%	23	Low
GSTM1	38-62%	3	Low
HRAS1	6%	1	Low
NAT2	56-62%	9	Low