

THE IMPACT OF KNOWN BREAST CANCER POLYGENES ON CRITICAL ILLNESS INSURANCE

BY CRAIG ADAMS, CATHERINE DONNELLY AND ANGUS MACDONALD

Department of Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical
Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

ABSTRACT

Genetic studies indicate that the inherited risk of breast cancer is mediated by the well-studied major genes BRCA1 and BRCA2, and a polygenic component, probably with many genes each making a small contribution. Recently, seven polygenes have been found (Pharoah *et al.*, 2008) contributing an estimated 3.6% of all familial risk (Easton *et al.*, 2007) This suggests that the polygenic component may involve well over 100 genetic loci.

We extrapolate these new results into a polygenic model with 147 genetic loci and simulate lifetimes of families to calculate the premium ratings appropriate for a family history of breast or ovarian cancer. We model the adverse selection costs arising from restricting the use of genetic test information in critical illness insurance underwriting in light of new European legislation banning the use of gender for insurance underwriting. In this setting we confirm the overall conclusion of Macdonald & McIvor (2009) that the polygene confers higher adverse selection risk than the BRCA genes. We establish that their three-gene polygenic model does not overly inflate the insurance costs attributable to a polygenic component of breast cancer risk under a model with 147 polygenes.

KEYWORDS

Genetics and Insurance; Adverse Selection; Family History; Breast Cancer; Polygene; BRCA1; BRCA2; Critical Illness

1. INTRODUCTION

Breast cancer (BC) is the most prevalent form of cancer in the world (Parkin, Bray, & Devesa, 2001) and a common cause of critical illness insurance claims. Established risk factors for BC include age, lifestyle factors, reproductive factors, and genetics (Washbrook, 2006). We consider the impact of the genetic risk on insurance products. Legal restrictions on the ability to use genetic information in insurance underwriting create an opportunity for individuals to select against an insurance company.

Rare inherited mutations in either of two genes, BRCA1 and BRCA2, confer greatly increased risk of BC and of ovarian cancer (OC). These are known as ‘single genes’ or ‘major genes’ because mutations in either are sufficient to increase the risk, in this case by affecting production of a DNA-repairing protein produced in the BC pathway (Tutt & Ashworth, 2002). The implications for life, critical illness and income protection insurance have been extensively studied (Macdonald, Waters, & Wekwete, 2003a,b; Gui *et al.*, 2006; Lu, Macdonald, & Waters, 2011; Lu *et al.*, 2011; Subramanian *et al.*, 1999; Lemaire *et al.*, 2000).

Current and future epidemiology is turning towards multifactorial models of genetic risk, in which variations in large numbers of genes, and environmental factors, interact to modify disease risk. Antoniou *et al.* (2002) proposed such a model for BC risk, known as a ‘polygenic’ model because it assumes: (a) that a large number of genes have two variants each; (b) that one variant of each gene increases risk and the other reduces risk; and (c) that the overall risk is the sum of the contributions from each gene (in a sense to be made precise later). The ‘high-risk’ varieties of each gene are not necessarily rare, but the additional risk they each confer is small compared with that conferred by BRCA1 and BRCA2 mutations. We refer to the collection of genetic loci in this model, and the variants at each, as a ‘polygene’.

In the absence of any known candidates for genetic loci contributing to the polygene, Antoniou *et al.* (2002) assumed that there were three, each contributing an effect of identical size independently of the others. Using the hypergeometric inheritance model of Lange (1997), they estimated the standard deviation of the distribution of the relative risk attributable to a BC polygene.

Macdonald & McIvor (2006) incorporated this polygenic model, along with models of BRCA1 and BRCA2 risk, into a Markov model of stand-alone critical illness (CI) insurance. They estimated premium

ratings conditional on either knowing an individual’s genotype (assuming genetic testing for the major genes and the polygene) or a family history of BC and OC. They found that the larger proportion of premium increases was the result of the polygene. However, they suggested that the heavy tails in the distribution of relative risks caused by there being only three genetic loci in the polygene could have skewed the results.

Seven single nucleotide polymorphisms (SNPs) that affect the onset of BC have subsequently been described (Easton *et al.*, 2007; Hunter *et al.*, 2007; Stacey *et al.*, 2007; Cox *et al.*, 2007). They are all candidates for the actual genetic loci in the polygene hypothesized by Antoniou *et al.* (2002). However, since in total they contribute about 3.6% of the heritable risk of BC (Easton *et al.*, 2007) and the two major genes another 25%, it is highly probable that the number of genetic loci involved is likely to be well over 100. We describe these developments in Section 2.1, and extend the findings to a plausible polygenic model accounting for 75% of the total heritable risk. We note that even since commencing work on this project using these seven SNPs, another eighteen loci have been identified with similar distributions, validating the assumptions we make in Section 2.2. The identification of the genetic loci that contribute to the risk of BC is progressing quickly and replication of results is already being sought for a further sixty-nine loci (Ghoussaini *et al.*, 2012).

In Section 3 we estimate the effect on stand-alone CI premiums of this polygenic model, given a family history of BC or OC, along the lines of Macdonald & McIvor (2006). We simulate the life histories of women within families. We find the distribution of genetic risk within families that develop a family history of BC or OC, and families that do not.

The European Court of Justice in March 2011 removed sex as a permissible insurance underwriting factor, from December 2012. This has particularly interesting implications for the pricing of BC risk, because men and women carry genotypes and develop family histories equally, but the risk of developing BC overwhelmingly affects women (male BC exists but is very rare). Thus men may disclose information affecting BC risk while being at virtually no risk of developing it themselves. We adapt our model to this new European context in Section 4. Additionally, in Section 5, we adapt the model to the terms of the moratorium in force in the United Kingdom which limits the use of genetic information in underwriting. This extends the insurance market model of Macdonald & McIvor (2006) to potential adverse selection in the light of recent epidemiology and in the new legal environment.

2. THE POLYGENIC MODEL

A polygenic model of inheritance exists where a trait is affected by variants of multiple genes. The trait of interest to us is the age-related risk of developing BC.

We assume that individual genes — BRCA1, BRCA2 and the loci contributing to the polygene — each show simple Mendelian inheritance. Thus a parent with a single mutated copy of either BRCA1 or BRCA2 (other risky major genotypes being rare enough to ignore) has a 50% chance of passing the mutation on to each offspring (male or female) they have. A parent with no such mutation (which we call the BRCA0 genotype) pass on only the BRCA0 genotype to offspring. Thus the increased BC risk associated with major gene mutations shows the same simple Mendelian inheritance as the major gene itself.

However BC risk conferred by the polygene does not show simple Mendelian inheritance. The phenotype associated with the polygene is the aggregate of the ‘phenotypes’ associated with each contributing genetic locus. Although these ‘phenotypes’ may be thought of as each obeying Mendel’s laws, they are effectively unobservable individually. Thus, a mother who has an average polygenic risk of BC may have daughters whose polygenic risk can range between very high and very low. Especially when the number of contributing loci is high, this may distort the pattern and usefulness of a family history of BC.

2.1 *Pharoah’s Model*

The model of Pharoah *et al.* (2008) combines multiplicatively the effects of seven known genetic loci to assign a total relative risk of BC. They are assumed to have no effect on OC risk. Since everyone has two functioning alleles of every gene (except those on the X and Y chromosomes), there are three possible combinations of alleles at each locus: no high-risk alleles; one low-risk allele and one high-risk allele; or two high-risk alleles. If the population prevalence of a high-risk allele at locus i is p_i , these three genotypes have population prevalences $(1 - p_i)^2$, $p_i(1 - p_i)$ and p_i^2 respectively. Table 1 sets out

Table 1: Per allele risk with 95% confidence intervals and population prevalences of genetic loci known to contribute to the polygenic risk of breast cancer (Pharoah *et al.*, 2008) with the approximate standard deviation of the per allele risk estimate.

Locus i	dbSNP Number	Per Allele Risk (95% Confidence Interval)	Population Prevalence, p_i	σ_i
1	rs2981582	1.26 (1.23, 1.30)	0.38	0.0179
2	rs3803662	1.20 (1.16, 1.24)	0.25	0.0200
3	rs889312	1.13 (1.10, 1.16)	0.28	0.0153
4	rs3817198	1.07 (1.04, 1.11)	0.30	0.0179
5	rs13281615	1.08 (1.05, 1.11)	0.40	0.0153
6	rs13387042	1.20 (1.14, 1.26)	0.50	0.0179
7	rs1053485	1.13 (1.18, 1.06)	0.86	0.0310

the estimated per allele risk and associated 95% confidence interval of the high-risk variant relative to the low-risk and population prevalences for each of these polygenes with their identification number in the SNP database¹, commonly referred to as ‘dbSNP Number’, as well as our own reference label, locus i for $i = 1, 2, \dots, 7$. Also given is a calculation of the standard deviation of the estimate, calculated as

$$\sigma_i = \frac{\text{Upper Confidence Interval} - \text{Lower Confidence Interval}}{2 \times 1.96}, \text{ for } i = 1, 2, \dots, 7. \quad (1)$$

2.2 Extending The Model

Easton *et al.* (2007) suggests that these seven loci account for 3.6% of familial risk, while the major genes BRCA1 and BRCA2 account for another 25%. Antoniou *et al.* (2001) concluded that a polygenic model is a better fit for the remaining familial risk than any model incorporating another major gene (a putative BRCA3, which has not been found despite intensive searching). We assume that the genetic loci which contributes to the polygene and have yet to be discovered, have characteristics broadly similar to the seven described above. Therefore we assume that there are 20 additional sets of seven loci with relative risk and associated confidence intervals to those of the set of seven known loci. Thus in total we have 147 loci with the 140 undiscovered loci labelled locus 8, 9, 10, \dots , 147. For convenience, we index the postulated loci such that loci 8, 15, \dots , 141 all have the same characteristics as locus 1; loci 9, 16, \dots , 142 have the same characteristics as locus 2, and so on. These additional 140 loci account for the remaining 71.4% of familial risk. Thus, our model is an extrapolation from the characteristics of known loci to complete the posited polygene.

The interaction of major gene mutations with genetic loci contributing to the polygene was not discussed in Pharoah *et al.* (2008). Further research has suggested where interactions may or may not exist (see Antoniou *et al.*, 2008), but this is not yet well established, so we have assumed that the effect of the polygene risk on a BRCA1 or BRCA2 mutation carrier is the same as that on a non-carrier. Since BRCA mutations are rare — Antoniou *et al.* (2002) estimated mutant allele frequencies of BRCA1 and BRCA2 as 0.00051 and 0.00068 respectively — this should not distort results greatly.

2.2.1 Notation

We introduce notation to describe genotypes, hazard rates and relative risks.

- (a) Let \mathcal{G} be the set of all possible polygenotypes. Assume that n genetic loci contribute to the polygene. Let \mathcal{G}_i be the set of all possible genotypes at the i th genetic locus. (For our model, $\mathcal{G}_i = \{0, 1, 2\}$ for all i would suffice.) Then the polygenotype in \mathcal{G} of a woman drawn at random from the population is a random variable denoted by $G = (G_1, G_2, \dots, G_n)$, with $G_i \in \mathcal{G}_i, i = 1, \dots, n$. Additionally, we denote by $g = (g_1, g_2, \dots, g_n)$ a realisation of G .

¹<http://www.ncbi.nlm.nih.gov/projects/SNP/>

- (b) Choose a starting age x low enough that no cases of BC have occurred. As noted in Antoniou *et al.* (2001), the proportions of lives in the population with each polygene mutation will change as higher-risk lives get BC and die earlier than lower-risk lives, lowering the average risk of survivors. The population prevalence of major genes and the polygene at age x will be that arising from Mendelian inheritance, with no survivorship effect yet.
- (c) Let ${}_t p_x^g$ be the probability that a woman free of BC at age x is still free of it age $x + t$, given she has genotype g .
- (d) Let p_g be the population prevalence at age x of polygenotype $g \in \mathcal{G}$.
- (e) Denote the age-dependent population onset rate of BC at age $x + t$ by $\mu^{BC}(x + t)$, and the polygenotype-specific onset rate at age $x + t$ by $\mu_g^{BC}(x + t)$ for polygenotype $g \in \mathcal{G}$.
- (f) Define $\lambda(x + t)$ to be the baseline onset rate of BC at age $x + t$. We choose our baseline genotype to be a woman with zero high-risk variants at each locus contributing to the polygene.
- (g) Define RR_{i,g_i} to be the relative risk associated with polygenotype $g_i \in \mathcal{G}_i$ at locus $i \in \{1, \dots, n\}$, and RR_g to be the relative risk for polygenotype $g \in \mathcal{G}$, assuming that the relative risks associated with genotypes at each contributing genetic locus are constant (so we have proportional hazards). The average relative risk within the population at age $x + t$ is denoted $\overline{RR}(x + t)$.
- (h) For the i th locus, define

$$\beta_i = \log(RR_{i,1}), \quad (2)$$

where $RR_{i,1}$ is the per allele relative risk of the high-risk variant of locus i , given in Table 1.

With our notation established, we can now set out the main definition of our model. The multiplicative model is defined by assuming that a woman with genotype g_i at the i th locus ($i = 1, \dots, n$) has onset rate of BC equal to:

$$\mu_g^{BC}(x + t) = \lambda(x + t)RR_g = \lambda(x + t)RR_{(g_1, \dots, g_n)} = \lambda(x + t) \prod_{i=1}^n RR_{i,g_i}. \quad (3)$$

In our simple model, we may define g_i to be the number of high-risk alleles carried at the i th locus. Then

$$\mu_g^{BC}(x + t) = \mu_{(g_1, \dots, g_n)}^{BC}(x + t) = \lambda(x + t) \exp(\beta_1 g_1 + \dots + \beta_n g_n), \quad (4)$$

which is a Cox-type proportional hazards model with the g_i as covariates.

2.2.2 Distribution of Relative Risk

Since each locus contributes three possible genotypes, and we have assumed that there are 147 loci, in total the polygene has 3^{147} (of the order of 10^{70}) variants. For comparison, the population of the world today is less than 10^{10} . We have modelled the 147 loci as 21 independent duplicates of the set of seven known loci. This means that many of the resulting 3^{147} polygenotypes have identical relative risks. However there are still too many to carry out computations by direct summation over all polygenotypes. We therefore find an approximate distribution for relative risk to allow the use of numerical methods to simplify the calculations.

Let RR denote the relative risk of an arbitrary polygene variant, relative to the polygene which has low-risk alleles at all loci. The distribution of $\log(RR)$ can be assumed to be approximately normal. Antoniou *et al.* (2002), using the hypothetical polygene with three loci, gave a mean of 1.29 and a 95% confidence interval of (1.096, 1.521) for the standard deviation of $\log(RR)$, which we denote $\sigma_{\log(RR)}$. Under our assumptions above, $\log(RR)$ has standard deviation $\sigma_{\log(RR)} = 1.20$, reasonably close to the simpler model (which was used in Macdonald & McIvor, 2006).

In this section, we assume that the genotype at each locus is denoted by the number of high-risk alleles denoted G_i , for $G_i \in \mathcal{G}_i = \{0, 1, 2\}$. Define p_i^* to be the population prevalence of the high-risk allele at the i th locus, so $G_i \sim \text{Bin}(2, p_i^*)$ for $i = 1, 2, \dots, 147$. Then the relative risk contributed from locus i , RR_i in a woman drawn at random from the population is a random variable, taking on the values $RR_{i,0}$, $RR_{i,1}$ and $RR_{i,2}$ with probabilities $(1 - p_i^*)^2$, $2p_i^*(1 - p_i^*)$ and p_i^{*2} respectively. G_1, G_2, \dots, G_{147}

are assumed to be independent random variables and for $j = 1, \dots, 7$, the random variable G_{j+7k} has the same distribution as G_j , for each $k = 0, 1, \dots, 20$.

Then the (random) log relative risk can be written as

$$\log RR_i = \beta_i G_i. \quad (5)$$

We assume that $p_{j+k}^* = p_j$ for $j = 1, 2, \dots, 7$ and $k = 0, 1, \dots, 20$ with the value of p_j given by Table 1. For $j = 1, \dots, 7$, let G_j^* be the random variable representing the total number of high-risk variants in the 21 analogues of locus j , in a woman drawn at random from the population, *i.e.*

$G_j^* = \sum_{k=0}^{20} G_{j+7k} \sim \text{Bin}(42, p_j^*)$. This distribution can be well approximated by a normal distribution: $G_j^* \sim \mathcal{N}(42p_j^*, 42p_j^*(1-p_j^*))$. Moreover, let $RR_{j,G_j^*}^*$ be the random variable representing the total relative risk contributed by the 21 analogues of that locus, in that same woman, *i.e.* $RR_{j,G_j^*}^* = \prod_{k=0}^{20} RR_{j+7k}$ and

$$\log RR_{j,G_j^*}^* = \sum_{k=0}^{20} \log RR_{j+7k} = \sum_{k=0}^{20} \beta_{j+7k} G_{j+7k} = \sum_{k=0}^{20} \beta_j G_{j+7k} = \beta_j G_j^*. \quad (6)$$

Hence the total relative risk RR_G of a woman drawn at random from the population has a distribution given by:

$$\log RR_G = \sum_{j=1}^7 \beta_j G_j^* \sim \mathcal{N} \left(\sum_{j=1}^7 42\beta_j p_j^*, \sum_{j=1}^7 42\beta_j^2 p_j^*(1-p_j^*) \right). \quad (7)$$

Hence, relative to polygenotype $g = (0, 0, \dots, 0)$, the polygene relative risk is approximately log-normally distributed with parameters $\mu_{\log RR} = 17.4289$ and $\sigma_{\log RR}^2 = 1.441$ at outset.

Figure 1 compares the quantiles of $\log(RR_G)$ obtained by sampling from the underlying binomial distributions for the $\log(RR_G)$, with a standard normal distribution. This shows visually that the normal distribution is a good fit to our assumed polygene model.

2.2.3 Baseline Rate of Onset of Breast Cancer

In equation 3, $\mu_g^{BC}(x+t)$ is defined in terms of baseline hazard rate, $\lambda(x+t)$. However, $\lambda(x+t)$ is not directly observable. What is observable is the population average onset rate, $\mu^{BC}(x+t)$, which embodies the survivorship effect mentioned in (b) of Section 2.2.1 above. Here we show how to calculate $\lambda(x+t)$ and constrain the average onset rate of BC for survivors at age $x+t$ in the model to be equal to the population onset rate. Noting that the mean relative risk at age $x+t$, $\overline{RR}(x+t)$, is given by:

$$\overline{RR}(x+t) = \frac{\sum_{g \in \mathcal{G}} p_g {}_t p_x^g RR_g}{\sum_{g \in \mathcal{G}} p_g {}_t p_x^g}, \quad (8)$$

we use the relation

$$\lambda(x+t) = \frac{\mu^{BC}(x+t)}{\overline{RR}(x+t)}, \quad (9)$$

To find $\lambda(x+t)$, we solve differential equations for ${}_t p_x^g$ for $g \in \mathcal{G}$

$$\frac{d}{dt} {}_t p_x^g = -\mu_g^{BC}(x+t) {}_t p_x^g = -\lambda(x+t) RR_g {}_t p_x^g = -\frac{\mu^{BC}(x+t)}{\overline{RR}(x+t)} {}_t p_x^g RR_g = -\frac{\mu^{BC}(x+t) \sum_{h \in \mathcal{G}} p_h {}_t p_x^h}{\sum_{h \in \mathcal{G}} p_h {}_t p_x^h RR_h} {}_t p_x^g RR_g. \quad (10)$$

Using the approximate distribution derived in Section 2.2.2, the sums over genotypes in equation 10 become integrals, *i.e.*

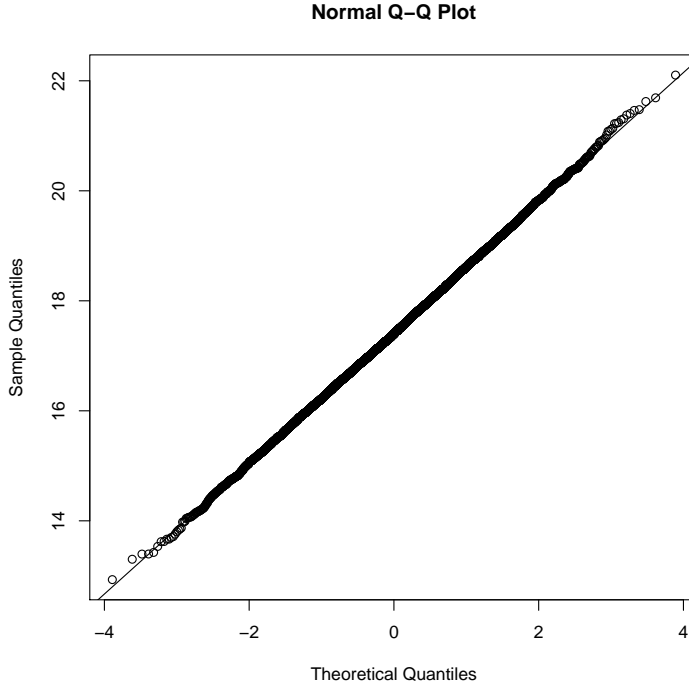


Figure 1: Q-Q plot of Log Relative Risk with 21 sets of 7 polygenes from sampling and a standard Normal distribution.

$$\frac{d}{dt} {}_t\tilde{p}_x^r \approx - \frac{\mu^{BC}(x+t) \int_{-\infty}^{\infty} f_{\log RR_G}(s) {}_t\tilde{p}_x^s ds}{\int_{-\infty}^{\infty} f_{\log RR_G}(s) {}_t\tilde{p}_x^s \exp(s) ds} {}_t\tilde{p}_x^r \exp(r). \quad (11)$$

where ${}_t\tilde{p}_x^r$ is the probability that a woman free of BC at age x is still free of BC at age $x+t$ given that the logarithm of her relative risk is r , for $-\infty < r < \infty$, and $f_{\log(RR_G)}(s)$ is the probability density function of $\log RR_G$. We calculated these integrals numerically using Simpson's Rule over the range $[-6.25, 6.25]$ with a step size of 2^{-6} .

The simultaneous differential equations were then solved using a fourth order Runge-Kutta method with step size 2^{-13} and the boundary conditions that ${}_0\tilde{p}_x^r = 1$ for all r .

3. COST OF FAMILY HISTORY

3.1 Insurance Model

In this section we consider women who develop a family history of BC or OC and the impact this has on premiums payable for insurance. This allows us to consider the appropriateness of the use of family history when underwriting in the absence of tests for either the major genes (MG) — BRCA1 and BRCA2 — or polygenes (PG).

The insurance contract of interest is stand-alone critical illness insurance with benefits payable on the occurrence of a defined critical illness, but not on earlier death. This type of product is of interest due to the high proportion of claims that can be directly attributed to BC.

Consistent with previous work, we define a family history to be one in which two or more first degree relatives have been diagnosed with BC or OC before age 50.

We construct a Markov model for the lifetime of an individual female as illustrated in Figure 2. To incorporate the major genes, BRCA1 and BRCA2, we augment the set of possible genotypes to

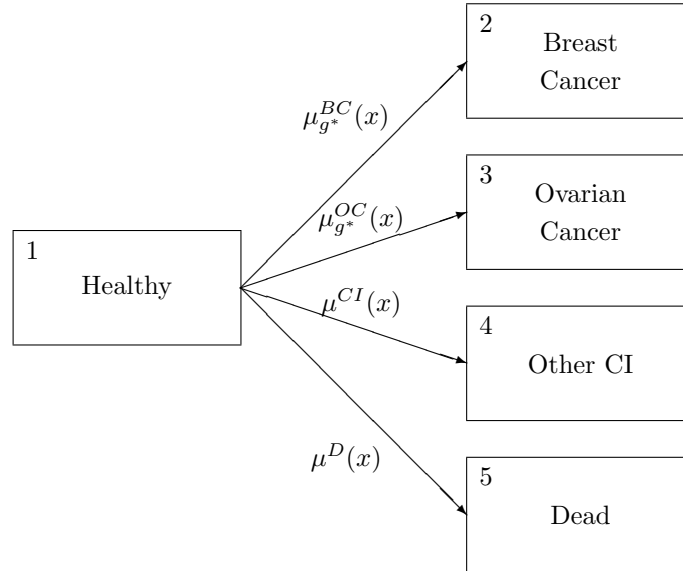


Figure 2: A model for critical illness insurance policyholder. Transition to ‘Dead’ or ‘Other CI’ states is at rates depending only on age x . The onset rate of OC, $\mu_g^{OC}(x)$, depends also on the BRCA genotype but is unaffected by polygenotype part of g , while the onset rate of BC, μ_g^{BC} , depends on both the polygenotype and BRCA genotype parts of g .

$\mathcal{G}^* = \{0, 1, 2\} \times \mathcal{G}$. Observed overall genotypes are thus $g^* = (g_0, g_1, \dots, g_{147})$ where $g_0 \in \{0, 1, 2\}$ is the major genotype corresponding to BRCA0, BRCA1, BRCA2 respectively. For consistency with previous work, transition intensities from ‘Healthy’ to ‘Dead’ and ‘Other CI’ are taken from Gutiérrez & Macdonald (2003). These include onset of all cancers so onset rates of BC and OC have been removed from them, using transition intensities from Macdonald, Waters, & Wekwete (2003a) as these were calculated from the same data (Cancer Registrations 1990–1992). BC and OC onset rates are those used in Macdonald & McIvor (2006), calculated from cancer registrations in England and Wales 1983–1987, for consistency with the adjustments for BRCA mutations estimated by Antoniou *et al.* (2002).

3.2 Premiums

Premium rates for each relative risk are calculated as continuously payable level net premiums by solving Thiele’s equations using a Runge-Kutta algorithm with step size 2^{-11} years and force of interest 0.05 *per annum*. A summary of these premiums as a percentage of those chargeable to a life with no BRCA mutation and $RR = \overline{RR}_x$ is shown in Table 2, allowing us to see how deviations from the population average risk at outset impacts premium rates.

First we examine premium rates for lives with no BRCA1 or BRCA2 mutation. Premiums for very low risk lives (BRCA0 and $\log RR < \log \overline{RR} - 1$) approach a limit of between 70-80% of premiums chargeable to a life with an average level of risk (BRCA0 and $\log RR = \log \overline{RR}$) depending on age and term. Premiums for very low risk lives can be attributed to the amount required to cover costs of ovarian cancer and other critical illness. However for even slightly elevated risk (BRCA0 and $\log RR = \log \overline{RR} + 1$) premiums should be loaded approximately 40%, reaching uninsurable rates ($> 400\%$ of unrated premium) for very high risk lives (BRCA0 and $\log RR = \log \overline{RR} + 3$), although the proportion of lives with such a high risk is small.

If we ignore the polygene effect by looking at $\log RR = \log \overline{RR}$, women with a BRCA mutation are uninsurable at most ages and terms. In our model which incorporates both a major gene and polygenes, a woman with a BRCA mutation but low polygenic relative risk becomes insurable. However in most of these low relative risk cases, the effect of the BRCA gene is still stronger than the decrease in BC risk from the polygene because the polygene does not change the high risk for ovarian cancer caused by BRCA mutations.

Table 2: Female premium rates as percentage of the premium rates for a life with no BRCA mutation and polygene relative risk $\log RR = \log(\overline{RR}_x)$.

BRCA		Age 30			Age 40		Age 50
		Term 10 %	Term 20 %	Term 30 %	Term 10 %	Term 20 %	Term 10 %
0	$\log(\overline{RR}_x) - 3$	76	73	77	71	77	81
	$\log(\overline{RR}_x) - 2$	78	76	79	74	79	83
	$\log(\overline{RR}_x) - 1$	84	82	85	81	85	87
	$\log(\overline{RR}_x)$	100	100	100	100	100	100
	$\log(\overline{RR}_x) + 1$	143	148	141	152	142	134
	$\log(\overline{RR}_x) + 2$	260	277	250	294	255	228
	$\log(\overline{RR}_x) + 3$	576	612	521	675	553	480
1	$\log(\overline{RR}_x) - 3$	113	240	208	327	242	173
	$\log(\overline{RR}_x) - 2$	164	278	232	359	261	181
	$\log(\overline{RR}_x) - 1$	303	381	300	445	313	204
	$\log(\overline{RR}_x)$	678	655	481	680	456	266
	$\log(\overline{RR}_x) + 1$	1675	1360	955	1305	852	433
	$\log(\overline{RR}_x) + 2$	4238	3040	2112	2929	1954	885
	$\log(\overline{RR}_x) + 3$	10301	6626	4598	6965	4877	2091
2	$\log(\overline{RR}_x) - 3$	106	97	115	90	117	140
	$\log(\overline{RR}_x) - 2$	144	128	141	119	141	161
	$\log(\overline{RR}_x) - 1$	246	213	210	195	204	217
	$\log(\overline{RR}_x)$	521	441	392	402	373	369
	$\log(\overline{RR}_x) + 1$	1260	1026	834	954	810	781
	$\log(\overline{RR}_x) + 2$	3184	2416	1787	2397	1865	1878
	$\log(\overline{RR}_x) + 3$	7885	5303	3699	5987	4317	4779

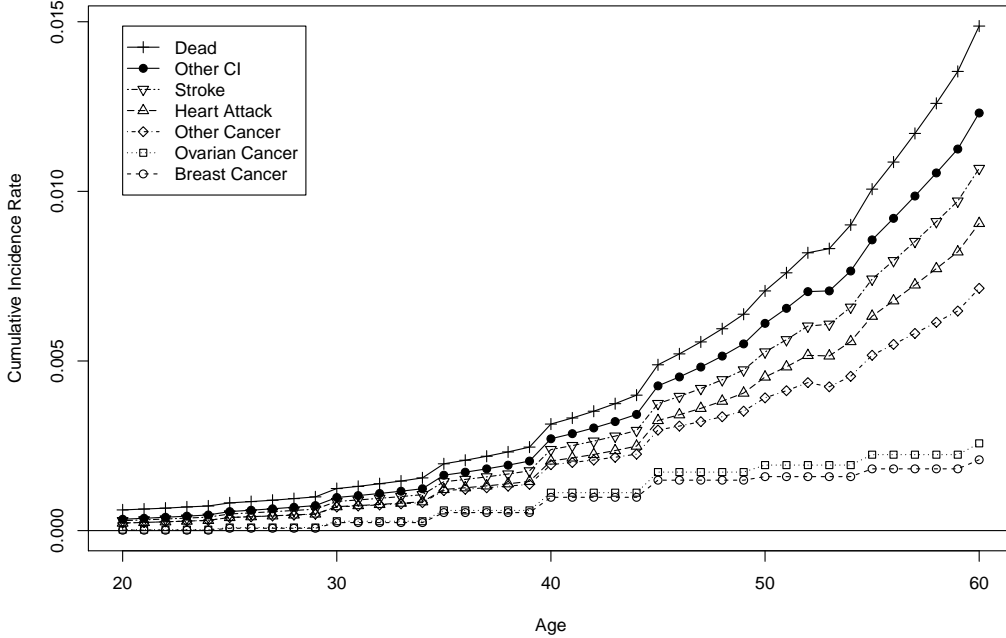


Figure 3: Stacked onset rates for selected critical illnesses and death of females.

The ratings due to polygene vary by age and term, the highest ratings for lives without a BRCA mutation being for 40 year olds with a 10 year term. The reason for this can be seen in Figure 3: BC is a significant proportion of the onset rates and this proportion reaches a peak between the ages of 40 and 50 so the polygene has the highest impact on policies with terms that include this age range. Its influence decreases as heart attack and other forms of cancer unaffected by the polygene reduce the weighting of BC in the makeup of premium rates.

For lives with a BRCA mutation, the highest ratings are earlier than the highest for lives without a BRCA mutation — 30 year olds with 10 year term compared to 40 year olds with 10 year term respectively — because the major genes have their strongest impact between ages 30 and 40, although BC is overwhelmingly dominant at all ages considered.

3.3 Simulation

Due to the large number of possible genotypes, and inheritance possibilities, it is not tractable to compute onset rates of family history analytically. Instead we simulate the future lifetimes of a large number of lives and their families.

The simulation model that we use is described fully in Macdonald & McIvor (2006). We summarise it here, and set out the change to inheritance methodology.

- (a) A family starts with 2 parents, whose polygenotype and BRCA genotype are randomly sampled according to the population distributions in Table 1.
- (b) The number of daughters is randomly sampled from a random variable with the probability mass function given in Table 3. The distribution is from Macdonald, Waters, & Wekwete (2003a). Daughters are the same age and born when the mother, assumed to be healthy, is age 30.
- (c) Each daughter inherits polygenes and BRCA genes from their parents independently of each other according to Mendel’s laws, acting at each major gene locus and each polygene locus.
- (d) Per-allele relative risk for locus i , for $i = 1, 2, \dots, 147$, is sampled from a normal distribution with standard deviation σ_i and mean equal to the point estimate listed in Table 1.

Table 3: Probability mass function of number of daughters in a family conditional on there being at least one daughter. Source: Macdonald, Waters, & Wekwete (2003a).

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

- (e) Full life histories of mother and daughters are simulated using their respective genotypes and treating each of the decrements in Figure 2 as independent competing risks.

As noted in Macdonald & McIvor (2006), this results in censorship whereby a life who moves to another critical illness state cannot have a subsequent cancer but the effect is minimal as only around 6% of lives have made such a move to other critical illness states by age 50, which is our cut-off age for family history.

The simulation of a large number of lives (within 15,000,000 families) allows us to observe the distribution of relative risk at any age x .

3.4 Distribution of Relative Risk by Underwriting Class

In order to calculate the premiums to charge an insured population, we find the distributions of relative risk and BRCA mutations for every age for lives with and without a family history.

For each life, at every age $x + t$, we know her genotype, whether she is healthy and which relatives, if any, have so far been diagnosed with BC or OC before age 50. We categorise this large quantity of data at each age by: BRCA genotype; $\log RR$; and presence of family history. Since $\log RR$ is a continuous quantity we discretise it over the range $[-6.25, 6.25]$ into 100 ‘bins’ of uniform length 0.125. For family history categorisation, we define two underwriting classes

- (a) ST — Unrated — Females without a family history.
(b) FH — Females who have developed a family history.

Denote the mean and standard deviation of \log relative risk at age x in ST by $\mu_{\log RR}^{ST}(x)$ and $\sigma_{\log RR}^{ST}(x)$ respectively and similarly for FH .

We ran the model 500 times to produce confidence intervals for the estimates of the means and standard deviations of $\log RR$ and to check convergence, results were compared to 100 simulations with little difference so we omit the comparison. The resulting discretised distributions for the average of our simulations are shown in Figure 4 for ST and FH at ages $x = 30, 40$ and 50 . (Note that Figure 4 includes male lives which we need in simulations in Section 4.)

The distribution of relative risk of BRCA0 lives within each underwriting class, remains roughly the same at each age despite higher risk lives leaving the class (by either getting cancer or developing a family history). There is very little difference in the spread of \log_{RR} in ST and in FH ; $\sigma_{\log_{RR}}^{FH}(x)/\sigma_{\log_{RR}}^{ST}(x)$ is close to 1. However, as can be seen from the plots and Table 4, there is a significant difference in the mean of \log_{RR} in each class of approximately 1.2. This results in average relative risk within FH being approximately 250% higher than within ST at ages 30, 40 and 50, making it a good proxy for genetic risk.

For lives with a BRCA mutation, Tables 6 and 8 show that the differences between the mean of \log_{RR} in each underwriting class are lower than that of BRCA0 and decreases with age. While there are insufficient lives with a family history at age 30 to draw solid conclusion, there are significant differences between FH and ST at ages 40 and 50. The spread of \log_{RR} in FH is slightly smaller than ST for BRCA2 at age 50. However at all other ages, and for BRCA1, there are no significant differences.

Table 4: Mean of log relative risk for female lives with no BRCA mutations in underwriting classes ST and FH .

Age (x)	$\mu_{\log RR}^{ST}(x)$	95% Confidence Interval	$\mu_{\log RR}^{FH}(x) - \mu_{\log RR}^{ST}(x)$	95% Confidence Interval
20	17.43	(17.25, 17.60)	0.874	(0.459, 1.423)
30	17.43	(17.25, 17.59)	1.179	(1.064, 1.347)
40	17.42	(17.25, 17.59)	1.196	(1.157, 1.243)
50	17.40	(17.23, 17.57)	1.112	(1.084, 1.140)

Table 5: Standard deviation of log relative risk for female lives with no BRCA mutations in underwriting classes ST and FH .

Age (x)	$\sigma_{\log RR}^{ST}(x)$	$\sigma_{\log RR}^{FH}(x)/\sigma_{\log RR}^{ST}(x)$	95% Confidence Interval
20	1.203	1.025	(0.792, 1.307)
30	1.203	1.047	(0.966, 1.130)
40	1.199	1.011	(0.989, 1.011)
50	1.188	1.004	(0.991, 1.019)

Table 6: Mean of log relative risk for female lives with BRCA1 mutations in underwriting classes ST and FH .

Age (x)	$\mu_{\log RR}^{ST}(x)$	95% Confidence Interval	$\mu_{\log RR}^{FH}(x) - \mu_{\log RR}^{ST}(x)$	95% Confidence Interval
20	17.43	(17.25, 17.60)	0.742	(-0.810, 2.676)
30	17.43	(17.25, 17.60)	0.898	(0.000, 2.003)
40	17.31	(17.14, 17.46)	0.874	(0.750, 1.040)
50	17.17	(17.02, 17.31)	0.555	(0.467, 0.651)

Table 7: Standard deviation of log relative risk for female lives with BRCA1 mutations in underwriting classes ST and FH .

Age (x)	$\sigma_{\log RR}^{ST}(x)$	$\sigma_{\log RR}^{FH}(x)/\sigma_{\log RR}^{ST}(x)$	95% Confidence Interval
20	1.203	0.336	(0.133, 1.085)
30	1.203	0.677	(0.242, 1.448)
40	1.142	0.946	(0.871, 1.023)
50	1.085	0.976	(0.926, 1.032)

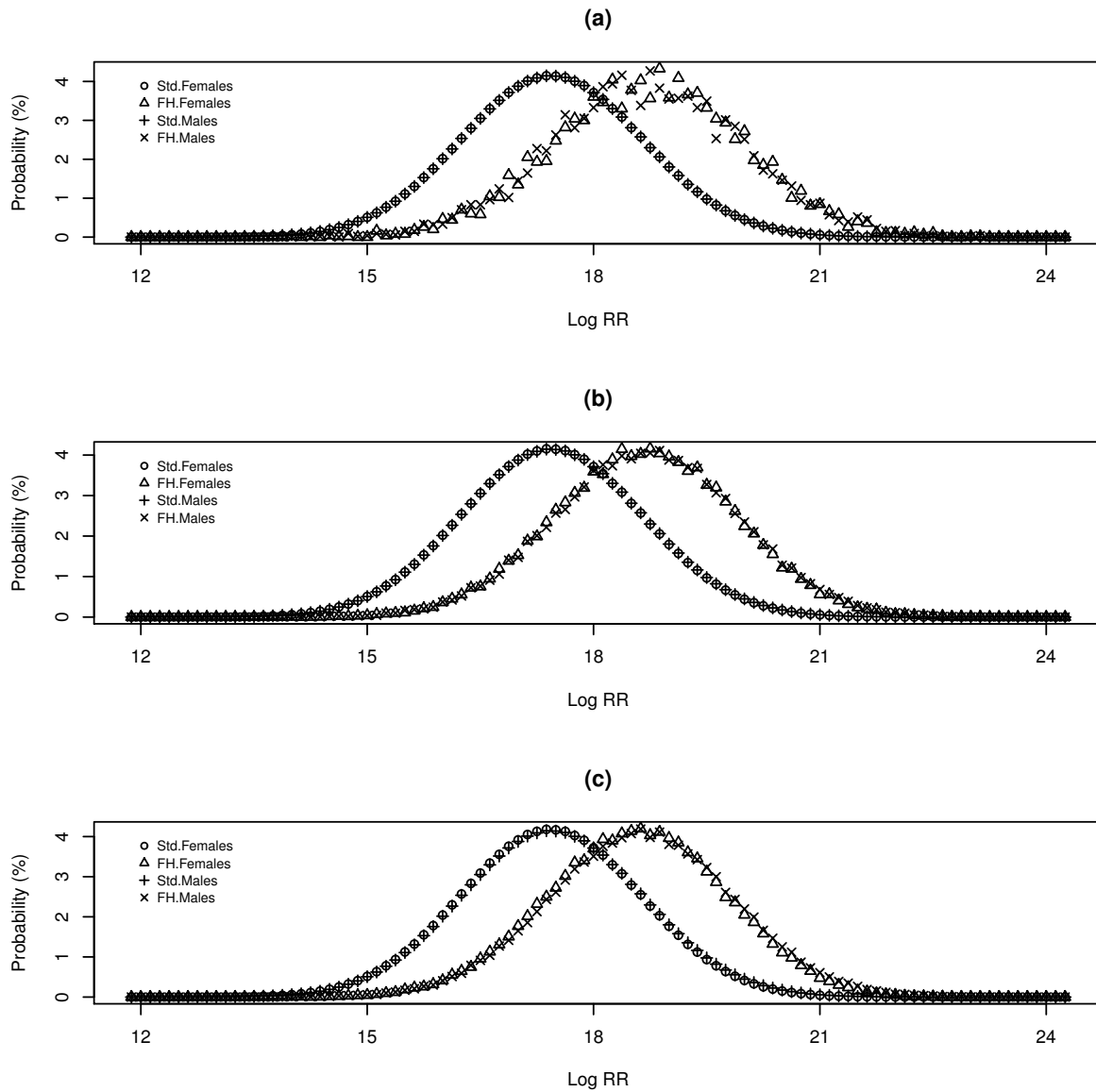


Figure 4: Distribution of relative risk for females and males at ages (a) 30, (b) 40, and (c) 50, averaged over 500 simulations.

Table 8: Mean of log relative risk for female lives with BRCA2 mutations in underwriting classes ST and FH .

Age (x)	$\mu_{\log RR}^{ST}(x)$	95% Confidence Interval	$\mu_{\log RR}^{FH}(x) - \mu_{\log RR}^{ST}(x)$	95% Confidence Interval
20	17.43	(17.26, 17.60)	0.560	(-0.241, 2.213)
30	17.43	(17.26, 17.60)	1.065	(0.001, 2.344)
40	17.34	(17.18, 17.49)	1.010	(0.876, 1.190)
50	17.20	(17.06, 17.35)	0.820	(0.729, 0.950)

Table 9: Standard deviation of log relative risk for female lives with BRCA2 mutations in underwriting classes ST and FH .

Age (x)	$\sigma_{\log RR}^{ST}(x)$	$\sigma_{\log RR}^{FH}(x)/\sigma_{\log RR}^{ST}(x)$	95% Confidence Interval
20	1.202	0.245	(0.086, 0.712)
30	1.202	0.425	(0.175, 1.376)
40	1.154	0.930	(0.830, 1.028)
50	1.098	0.905	(0.847, 0.977)

The proportions of lives with a BRCA mutation in each of the underwriting classes is also significantly different, as can be seen in Table 10.

3.5 Ratings for Presence of Family History

With these distributions of polygene relative risk and of major gene mutations, we will calculate the level premiums that would cover the cost of CI benefits payable assuming underwriters may use family history as a factor. We introduce notation to be used in this section. Suppose a life aged x has applied for a policy with term n years.

- $f_{x,\log RR,G_0|FH}(r,g)$ is the joint probability density function for $\log RR$ and major genotype, G_0 , conditional on a life aged x having developed a family history of BC or OC. For succinctness, this is abbreviated to $f_x^{FH}(r,g)$. A similar expression is used for ST .
- $A_{r,g,x:\overline{n}}$ is the expected present value of 1 paid on transition to a state other than Healthy or Dead for a life with relative risk e^r and major genotype g currently aged x .
- $a_{r,g,x:\overline{n}}$ is the expected present value of an annuity of 1 paid continuously while in the Healthy state for a life with relative risk e^r and major genotype g currently aged x .
- $\Pi_{x:\overline{n}}^{ST}$ is the level premium for a life without a family history.
- $\Pi_{x:\overline{n}}^{FH}$ is the level premium for a life with a family history.

Table 10: Proportions of lives with a BRCA mutation.

Age	ST		FH	
	BRCA1	BRCA2	BRCA1	BRCA2
30	0.0020	0.0027	0.0125	0.0112
40	0.0019	0.0025	0.0639	0.0518
50	0.0015	0.0023	0.0400	0.0293

Table 11: Distribution of rating for presence of family history.

Age	Term	Mean	95% Confidence Interval	Standard Deviation
30	10	1.927	(1.624, 2.252)	0.1661
	20	1.873	(1.627, 2.126)	0.1262
	30	1.685	(1.504, 1.869)	0.0930
40	10	2.473	(2.354, 2.594)	0.0596
	20	2.020	(1.932, 2.103)	0.0419
50	10	1.586	(1.548, 1.625)	0.0204

Table 12: Level net premiums for females with a family history of BC or OC as a percentage of those for females without a family history, averaged over simulations. Also shown are equivalent results from Macdonald & McIvor (2006). Note MG results use P+MG for experience basis but MG only for pricing basis.

Study	Genetic Model	Age 30			Age 40		Age 50
		10 years %	20 years %	30 years %	10 years %	20 years %	10 years %
Ours	P+MG	192.7	187.3	168.5	247.3	202.0	158.6
	MG	110.1	108.4	106.1	149.1	132.8	113.4
M&M (2006)	P+MG	444.0	341.0	274.7	244.2	207.4	170.6
	MG	137.5	132.0	122.7	112.9	108.9	102.8

The appropriate level premium for lives that have presented a family history is calculated as

$$\Pi_{x:\overline{n}}^{FH} = \frac{\sum_{g=0}^2 \left[\int_{-\infty}^{\infty} f_x^{FH}(r, g) A_{r, g, x:\overline{n}} dr \right]}{\sum_{g=0}^2 \left[\int_{-\infty}^{\infty} f_x^{FH}(r, g) a_{r, g, x:\overline{n}} dr \right]} \quad (12)$$

and a similar expression can be written down for $\Pi_{x:\overline{n}}^{ST}$.

We use the estimated distributions based on each of our 100 simulations to calculate estimated ratings and show how uncertainty may cause them to vary. Results of this are shown in Table 11 and Figure 5. They show a narrow confidence interval for age 50 at entry. However younger ages at entry have a broader interval, up to 65% of the unrated premium.

The approach taken by Macdonald & McIvor (2006) was to compare weighted premiums for lives with a family history (as calculated above) to the premiums appropriate for a life with BRCA0 and polygene relative risk $RR = 1$. However, the average polygene risk within the unrated pool will be somewhat higher — they used a binomial distribution for the log risk, centred at 0 — which increases the average premium for lives without a family history, offsetting some of the extra payable for the lives with a family history. In a reconciliation exercise, this was found to be around 40%.

Allowing for this change, and comparing our average ratings in Table 12, the results from the two polygene models do show some differences at ages at entry 40 and 50 of 30–40% while age 30 is far greater at all terms by 100–250%. Macdonald & McIvor ran their model only once and at age 30 they had 91 lives with a family history, 9 of which had a BRCA mutation as well as polygene relative risk greater than 1, resulting in large premium ratings. It is therefore possible that the difference between our results and theirs at age 30 is down to sampling. In either case, the largest part of rating is attributable to the polygenes.

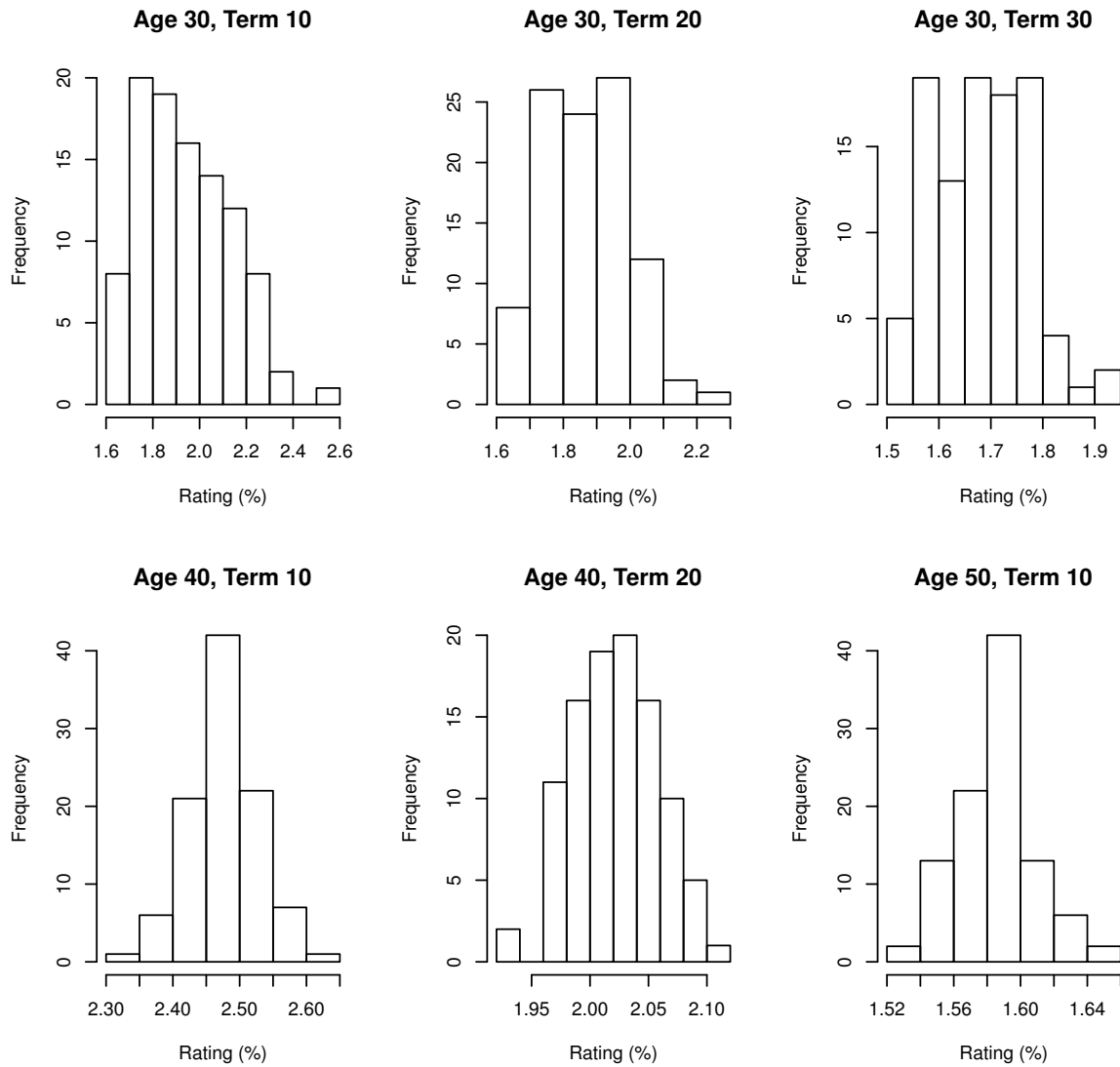


Figure 5: Histograms of ratings for presence of family history as a percentage of the premium charged to lives without a family history.

Table 13: Probability mass function of the number of children in a family. The probability that a particular child is born female is $1/2.06$.

Number of Children	Probability
0	0.090
1	0.130
2	0.400
3	0.200
4	0.137
5	0.031
6	0.009
7	0.003

Table 14: Premium rates for males as a percentage of premium rates for females with $\log RR = \log \overline{RR}$ and no BRCA mutation.

Age 30			Age 40		Age 50
Term 10	Term 20	Term 30	Term 10	Term 20	Term 10
74%	88%	104%	97%	113%	128%

4. EFFECT OF THE ‘TEST-ACHATS’ EUROPEAN COURT OF JUSTICE RULING

In March 2011 the European Court of Justice ruled, in a case brought by Belgian consumer organisation Test-Achats, that pricing with sex as a rating factor is contrary to the principles of the European Union and would not be permitted from 22nd December 2012, (see European Court of Justice, 2011). We assume that insurers’ response will be to charge premiums based on aggregate risk on an assumed business mix. It is not clear whether this extends to underwriting based on justifiable biological risk and it is possible that this would need to be tested in court. For example it is not clear whether or not a man, who has a family history of BC among his female relatives should be given the same rating his sister would be given. There are two possibilities:

- (a) Females receive a rating reflecting the average risk of a female with a family history and male lives are not rated.
- (b) Males and females are rated equally for a family history of BC or OC based on aggregate risk.

The model described above needs to be adjusted to include male lives. Each family is randomly assigned a number of children following the distribution of of family size and sex given in Macdonald, Waters, & Wekwete (2003a) (see Table 13). We define the probabilities a newborn is female or male as $p_f = 1/2.06$ and $p_m = 1.06/2.06$ respectively.

Research by Tai *et al.* (2007) shows an increased risk of male BC due to BRCA mutations, particularly BRCA2. However, detailed research is limited, partly due to the rarity of BC in males. BRCA mutations have also been linked to prostate cancer (see Narod *et al.*, 2008) but this is not usually a severe form of cancer and diagnosis is often missed until later in life. Consequently it would not be particularly helpful in identifying potential BRCA carriers from family history. We therefore perform simulations and calculate male premiums ignoring BRCA mutation in males except for transmission from father to daughter. Transition rates for males are those given in Gutiérrez & Macdonald (2003).

Table 14 gives the premium rates as a percentage of those for a female with $\log RR = \log \overline{RR}$ and no BRCA mutation.

The percentages of females and males who develop a family history in our simulations are shown in Table 15. At each age there is no significant difference between the percentages of females with a family

Table 15: Percentage of healthy lives with a family history of breast or ovarian cancer shown by sex and age.

Age	Females			Males		
	Mean	95% Confidence Interval	Standard Deviation	Mean	95% Confidence Interval	Standard Deviation
20	0.0002	(0.0001, 0.0003)	0.00005	0.0002	(0.0001, 0.0003)	0.00005
30	0.0020	(0.0015, 0.0028)	0.00034	0.0020	(0.0015, 0.0028)	0.00032
40	0.0220	(0.0169, 0.0285)	0.00334	0.0226	(0.0170, 0.0294)	0.00356
50	0.0880	(0.0680, 0.1117)	0.01290	0.0940	(0.0717, 0.1202)	0.01407

history of BC or OC and those of males.

To distinguish between males and females we introduce further notation

- (a) $f_x(r, g)$ is the joint distribution function for $\log RR$ and and major genotype, G_0 .
- (b) $f_{x, FH}$ and $m_{x, FH}$ are the probabilities that a female and male respectively, develop a family history of BC or OC by age x .
- (c) $A_{x:\bar{n}}^F$ and $A_{x:\bar{n}}^M$ are the expected present values of 1 paid to a life aged x in underwriting class ST immediately on transition to a claim state for females and males respectively for a term of n years.
- (d) $a_{x:\bar{t}}^F$ and $a_{x:\bar{t}}^M$ are the expected present values of a continuously payable annuity at rate 1 *per annum* while in an insured state, to a life currently aged x and in underwriting class ST for females and males respectively for a term of n years.
- (e) $A_{r,g,x:\bar{n}}^F$ and $a_{r,g,x:\bar{n}}^F$ are the female log relative risk and major genotype specific assurance and annuity functions defined above in Section 3.5.

For calculation of unisex premiums we set an even split between the sexes at each age of our insured population: $P(\text{Male}) = P(\text{Female}) = 0.5$.

First we calculate the expected present value of benefits for a female in ST aged x for term t as

$$A_{x:\bar{t}}^F = \sum_{g=0}^2 \left[\int_{-\infty}^{\infty} f_x^{ST}(r, g) A_{r,g,x:\bar{t}}^F dr \right], \quad (13)$$

and the expected present value of her continuously payable annuity as

$$a_{x:\bar{t}}^F = \sum_{g=0}^2 \left[\int_{-\infty}^{\infty} f_x^{ST}(r, g) a_{r,g,x:\bar{t}}^F dr \right]. \quad (14)$$

We want to rate lives in FH for only the extra risk they bring to the pool due to the higher onset rates of BC and OC females in the class will experience. It is then necessary to consider them as paying the premium $\Pi_{x:\bar{t}}^{ST}$ plus an additional sum. They are therefore included among the proportion of females when calculating our unisex premiums for underwriting class, ST , as

$$\Pi_{x:\bar{t}}^{ST} = \frac{A_{x:\bar{t}}^F P(\text{Female}) + A_{x:\bar{t}}^M P(\text{Male})}{a_{x:\bar{t}}^F P(\text{Female}) + a_{x:\bar{t}}^M P(\text{Male})}. \quad (15)$$

Note that setting the proportion of males, $P(\text{Male}) = 0$ results in the premiums calculated above in equation 12.

Treating females from FH as if they were in ST , who are on average lower risk, will create a shortfall in income which must then be spread across the rated business.

4.1 Females receive a rating, Males do not

Table 16: Distribution of ratings applicable only to females for a presence of family history where unrated premium is unisex.

Age	Term	Mean	95% Confidence Interval	Standard Deviation
30	10	2.076	(1.720, 2.473)	0.1983
	20	1.934	(1.666, 2.220)	0.1415
	30	1.674	(1.492, 1.863)	0.0961
40	10	2.504	(2.362, 2.657)	0.0741
	20	1.960	(1.868, 2.065)	0.0484
50	10	1.515	(1.473, 1.563)	0.0233

The simplest of the two underwriting possibilities recognises that the presence of a family history of BC or OC does not suggest a high risk for male lives and allows all males to belong to class ST . Males and females are charged the same unisex premium for the ‘standard’ risk. Females who have a family history pay an extra premium for the increased risk they pose. Thus the total premium for FH lives, $\Pi_{x:\bar{t}}^{FH}$, is

$$\Pi_{x:\bar{t}}^{FH} = \frac{P(\text{Female}) \sum_{g=0}^2 \int_{-\infty}^{\infty} f_x(r, g) \left(A_{r,g,x:\bar{t}}^F - \Pi_{x:\bar{t}}^{ST} a_{r,g,x:\bar{t}}^F \right) dr + P(\text{Male}) \left(A_{x:\bar{t}}^M - \Pi_{x:\bar{t}}^{ST} a_{x:\bar{t}}^M \right)}{f_{x,FH} P(\text{Female}) \sum_{g=0}^2 \int_{-\infty}^{\infty} f_x^{FH}(r, g) a_{r,g,x:\bar{t}}^F dr} + \Pi_{x:\bar{t}}^{ST}. \quad (16)$$

In comparing Tables 11 and 16, the effect of unisex premiums is to increase ratings slightly at younger ages where male CI costs are lower than females — reducing the premiums that females pay — and to decrease ratings slightly at older ages where males have high heart attack and stroke rates (see Figure 6) — increasing the premiums that females pay.

4.2 Females and Males receive a rating

If rating females but not males for a family history of BC or OC is seen as sex discrimination, the shortfall in unrated premium will be spread over males also, *i.e.* males with family history belong to underwriting class FH . Our rated premium becomes more general:

$$\Pi_{x:\bar{t}}^{FH} = \frac{P(\text{Female}) \sum_{g=0}^2 \int_{-\infty}^{\infty} f_x(r, g) \left(A_{r,g,x:\bar{t}}^F - \Pi_{x:\bar{t}}^{ST} a_{r,g,x:\bar{t}}^F \right) dr + P(\text{Male}) \left(A_{x:\bar{t}}^M - \Pi_{x:\bar{t}}^{ST} a_{x:\bar{t}}^M \right)}{f_{x,FH} P(\text{Female}) \sum_{g=0}^2 \int_{-\infty}^{\infty} f_x^{FH}(r, g) a_{r,g,x:\bar{t}}^F dr + m_{x,FH} P(\text{Male}) a_{x:\bar{t}}^M} + \Pi_{x:\bar{t}}^{ST}. \quad (17)$$

Since the extra cost of a high proportion of high-risk female lives in the FH underwriting class is spread over approximately twice the number of lives than when only females are rated, it is not surprising that the ratings and standard deviations in Table 17 are approximately half those in Table 16.

5. CALCULATING ADVERSE SELECTION IN AN INSURANCE MARKET

With the identification of these seven loci, the possibility of testing for an individual’s overall BC risk is one step closer. If a life has a pre-symptomatic genetic test, then under the UK’s moratorium on the use of genetic information in underwriting, the life is not required to divulge these test results. This asymmetry of information presents the life with an opportunity to change their buying behaviour. Lives

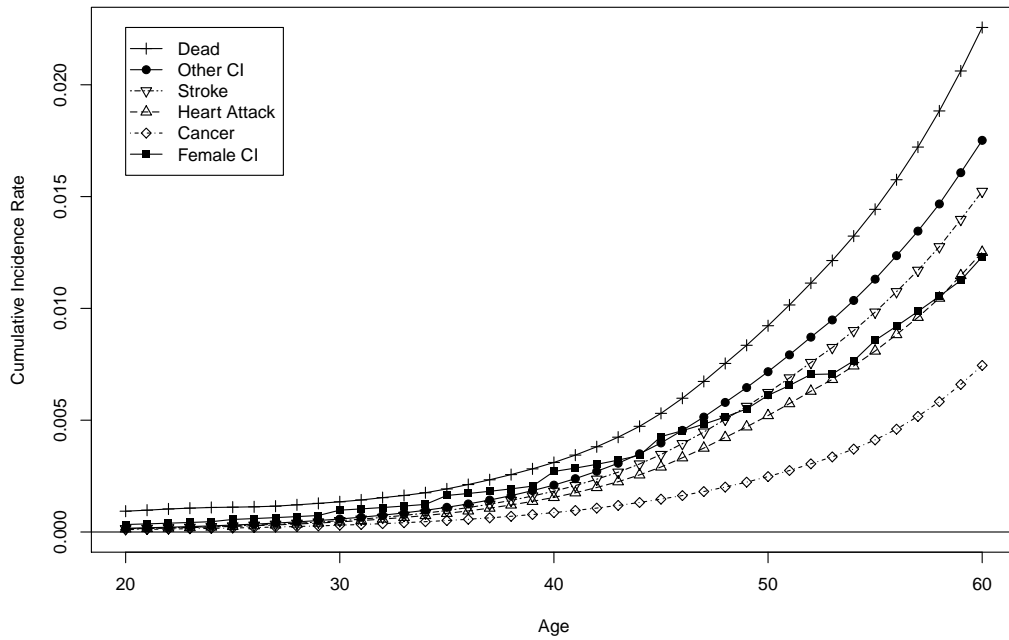


Figure 6: Stacked onset rates for selected critical illnesses and death in males. The population average female CI rate is overlaid for comparison.

Table 17: Distribution of rating applicable to male and female lives for presence of family history based on a unisex standard premium.

Age	Term	Mean	95% Confidence Interval	Standard Deviation
30	10	1.538	(1.364, 1.740)	0.1010
	20	1.465	(1.334, 1.599)	0.0719
	30	1.334	(1.243, 1.422)	0.0487
40	10	1.730	(1.659, 1.793)	0.0355
	20	1.463	(1.418, 1.507)	0.0227
50	10	1.248	(1.229, 1.269)	0.0106

who see themselves as low-risk may buy less insurance, while lives who see themselves as high-risk may buy more insurance. Both are forms of adverse selection.

Researchers have found adverse selection in insurance markets where individuals have private information: Brown (1992) in health insurance; Finkelstein & Poterba (2002) in annuities markets; and Finkelstein & McGarry (2006) in a market for long-term care. However, Zick *et al.* (2000) suggest that women who have tested positive for BRCA1 mutations are not more likely to purchase more life cover than untested women. Neither are women who have tested negative for BRCA1 less likely to purchase life cover than untested women. They suggest that their follow-up period is too short since many participants were still dealing with the health impacts of receiving a positive test result.

In this section we look at the potential cost of adverse selection based on the results of genetic tests.

We set up a model for an insurance market and parameterise it appropriately. The likely buying patterns and testing arrangements are uncertain so we investigate how the cost might change in response to these factors.

5.1 Market Model

We expand our insurance model to an insurance market, incorporating buying behaviour that may change due to the result of genetic testing. The product of interest is again stand-alone critical illness. To ensure the model is Markov, the premium is payable continuously as a current risk premium instead of a level premium. This risk premium at time t is calculated for each sex as the weighted average cost of expected claims arising in $(t, t + dt)$, with weights equal to occupancy probabilities in insured states assuming no adverse selection — all lives purchase insurance at the same standard rate. To conform to the EU legislation, these gender-specific risk premiums will have to be averaged across the two sexes with the weights for each being the overall probability a male and female is insured respectively, calculated assuming adverse selection to take place.

Since the premiums payable assume a lower-than-experienced rate of claim, a loss will arise. This loss is the cost of adverse selection and is calculated as the difference between outgo and premium income and expressed as a percentage of premium income received with adverse selection present, to show how much all premiums must increase as a result of the insurer not seeing test results.

All lives are assumed to be 20 years old, healthy, untested and uninsured at the start of our modelling. The model is illustrated in Figures 7 and 8.

Different scenarios are modelled to show the effects of uncertain parameters as follows:

- (a) The rates of going into a tested state are those used by Macdonald & McIvor (2009) which they based on Ropka *et al.* (2006)'s estimate of 59% uptake: High with rate 0.08916 *per annum*; Medium with rate 0.04458 *per annum*; or Low with rate 0.02972 *per annum*. All testing is performed on lives between ages 20 and 40.
- (b) Market size is at two levels represented by the rate at which a standard life enters the insured state: a Large market with rate 0.05 *per annum* or Small market with 0.01 *per annum*.
- (c) After receiving test results, a life may change their buying behaviour. Low risk lives either buy at the same standard rate, half the standard rate or do not buy at all. High risk lives, regardless of the size of market, will buy at rate 0.25 *per annum* which we take to represent a 'Severe' level of adverse selection; or at rate 0.1 *per annum* which we consider to be 'Moderate' adverse selection. A life is deemed to be high risk if their log relative risk is higher than the log of the population average relative risk plus the threshold in the scenario (see below). They are low risk if their risk is lower than the log of the population average relative risk minus the threshold.

The threshold of relative risk serves to classify lives into three categories of relative risk. It is varied to show the effect of how different from the population a life needs to see herself as being before she considers herself to be high risk or low risk *i.e.* whether the product represents high or low value to her. We limit ourselves to using high-risk and low-risk thresholds that are symmetric around the log of population average relative risk. However, in reality, the distance from this average that is needed to change behaviour may differ for low-risk and high-risk lives. We choose average relative risk as a central point because the premium charged will be approximately equal to the appropriate premium for a BRCA0 life with average relative risk. appropriate premium for a BRCA0 life with average relative risk When tested lives consider whether they are getting a cheap or expensive rate, it will be with reference to this

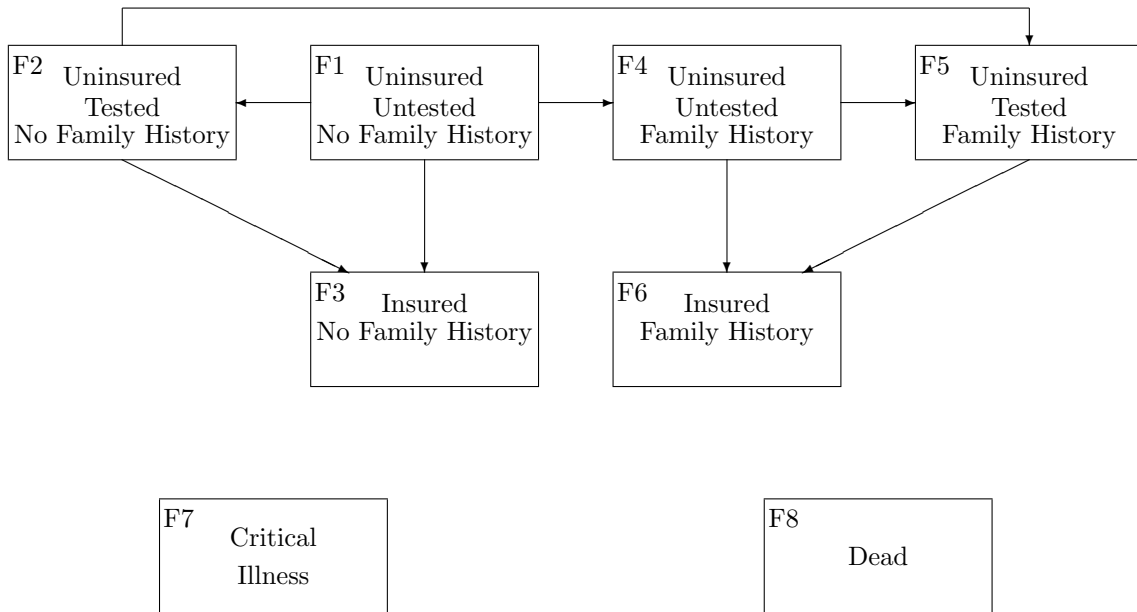


Figure 7: A model of female insurance states for an insurance market where genetic testing may be available before and after family history at different rates. The arrows to Dead and Critical Illness states are omitted but these may be entered from any other state.

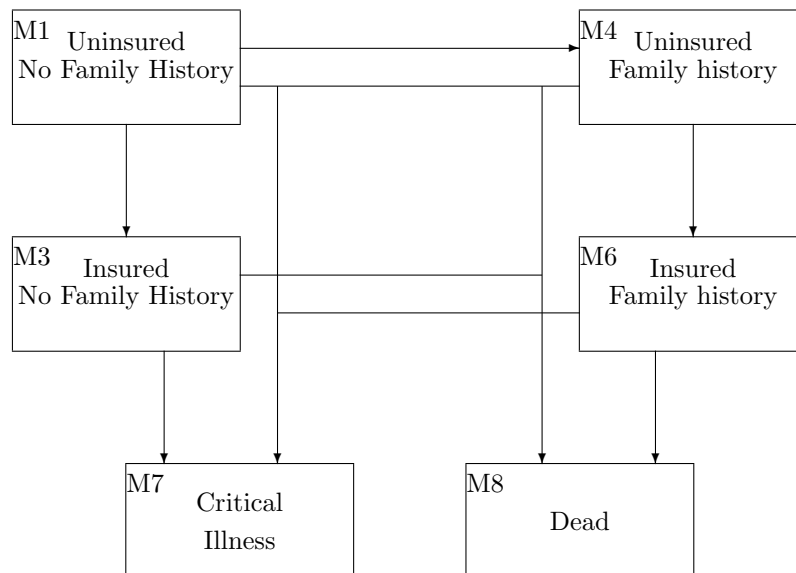


Figure 8: A model of male insurance states in an insurance market. Depending on legislation, lives in state state M6 will belong to either underwriting class *ST* or *FH*.

Table 18: Proportion of lives at age 20 considered Low or High Risk at each threshold point.

Threshold	0.0	0.5	1.0	1.5	2.0	2.5	3.0
Low Risk (%)	72.8	57.6	41.1	26.0	14.5	7.0	2.9
High Risk (%)	27.2	15.3	7.5	3.2	1.1	0.4	0.1

average premium. However, this will create an imbalance between the proportions of high and low risk lives (see Table 18). Macdonald & McIvor (2009) used the mode of their distribution as the centre.

5.2 Family History Onset Rate

The market model includes the onset of family history as an event. This intensity was parameterised by fitting a statistical model to the onset rate derived from the results of the simulations in Section 3 above and calculating

$$\text{Onset Rate} = \frac{\text{New family history cases in period}}{\text{Exposed to risk}}. \quad (18)$$

There is a large difference where a parent carries a BRCA mutation so these are considered separately. As there was no statistically significant difference between sexes, we ignore sex as a factor.

Denote the BRCA0 specific onset rate of family history as $\mu_{\text{BRCA0}}^H(x, r)$ for a life aged x with log $RR = r$ and similarly for BRCA1 and BRCA2.

As can be seen in Figures 9, 10 and 11, the dependence on the individual's log polygene risk is strongly linear.

Family history cannot arise after the life is age 50 as the definition is limited to cancer before age 50 and all lives are assumed to be born at the same time. Fitting linear models with interactions between age and risk yields formulae for the family history onset rates. For BRCA0 this results in the following:

$$\mu_{\text{BRCA0}}^H(x, r) = \begin{cases} \exp(-21 - 0.4696x + 0.0081x^2 + 0.0463xr - 0.0006x^2r), & \text{if } x < 50 \\ 0, & \text{if } x \geq 50. \end{cases} \quad (19)$$

Since the BRCA mutations are only assumed to have an effect from age 30 onwards, these rates are fitted in two stages, $x \leq 30$ and $30 < x < 50$. We pool the lives from all families where a BRCA mutation is present for $x \leq 30$ due to sparse data:

$$\mu_{\text{BRCA1}}^H(x, r) = \begin{cases} \exp(-12.44512 - 4.44898x + 0.2486xr), & \text{if } x \leq 30 \\ \exp(-37.63246 + 4.34126x - 0.2344xr + 1.5182r), & \text{if } 30 < x < 50 \\ 0, & \text{if } x \geq 50 \end{cases} \quad (20)$$

and

$$\mu_{\text{BRCA2}}^H(x, r) = \begin{cases} \exp(-12.44512 - 4.44898x + 0.2486xr), & \text{if } x \leq 30 \\ \exp(-31.44989 + 0.20735x - 0.011xr + 1.17576r), & \text{if } 30 < x < 50 \\ 0, & \text{if } x \geq 50. \end{cases} \quad (21)$$

5.3 Cost of adverse selection

The cost of adverse selection may be expressed as a percentage of premium income and defined as

$$\frac{\text{E(PV Benefits|Adverse Selection)} - \text{E(PV Premium Income|Adverse Selection)}}{\text{E(PV Premium Income|Adverse Selection)}}, \quad (22)$$

with a risk premium rate calculated assuming no adverse selection occurs. This gives a value for how much premium rates of all lives must increase as a result of adverse selection being present in the market.

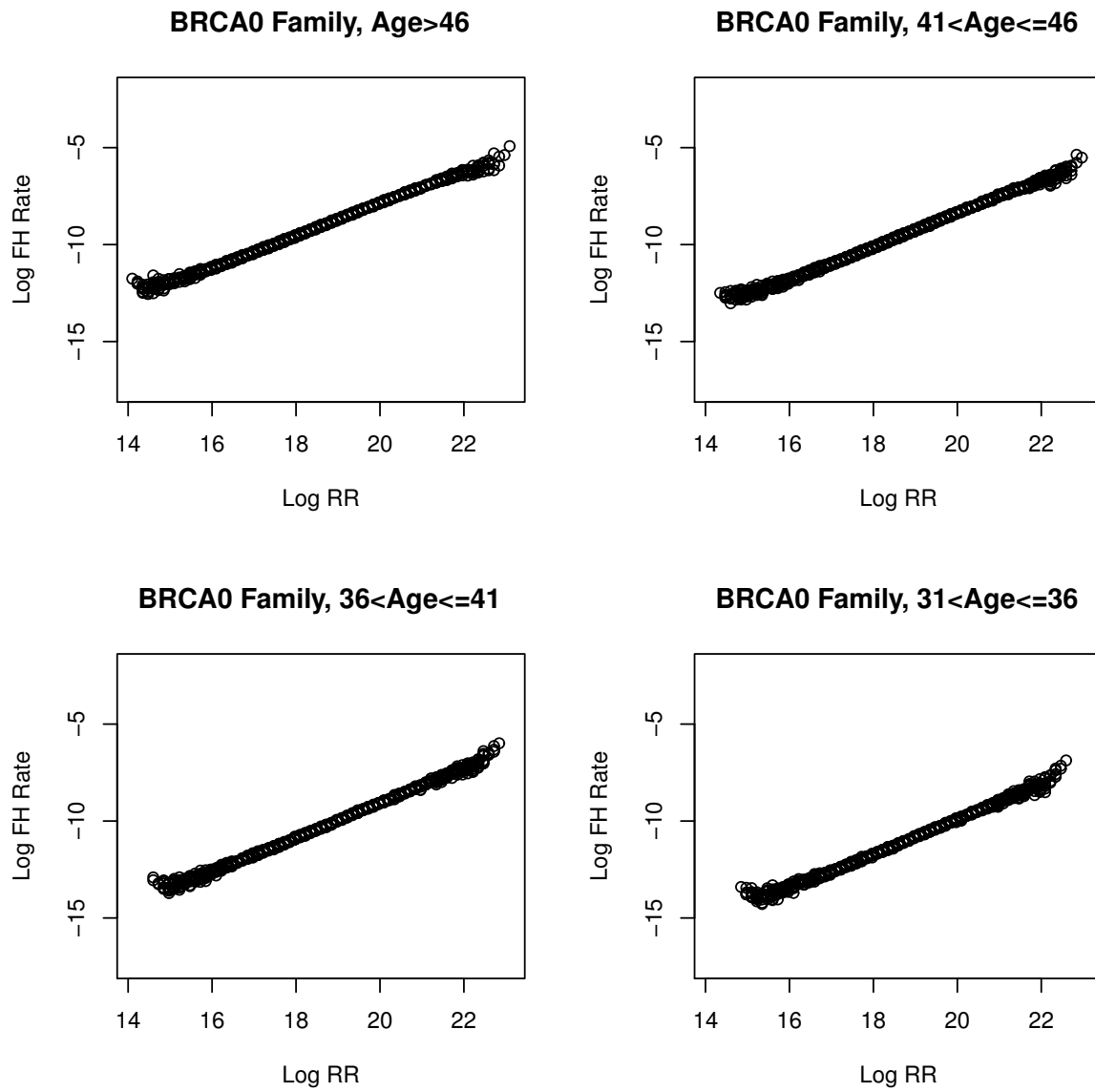


Figure 9: Logarithm of family history onset rates for families with no BRCA mutation in either parent. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.

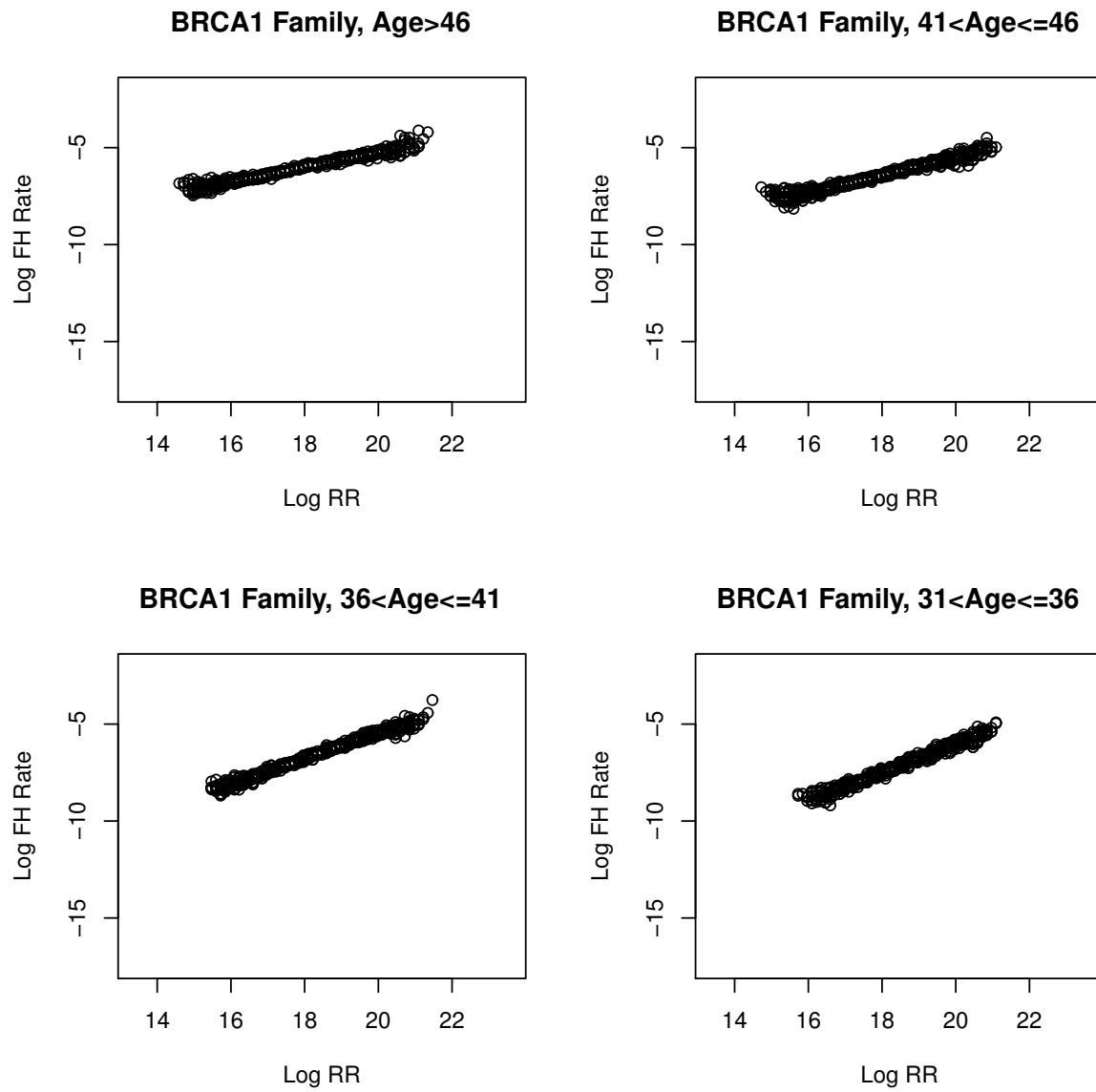


Figure 10: Logarithm of family history onset rates for families where a parent has a BRCA1 mutation. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.

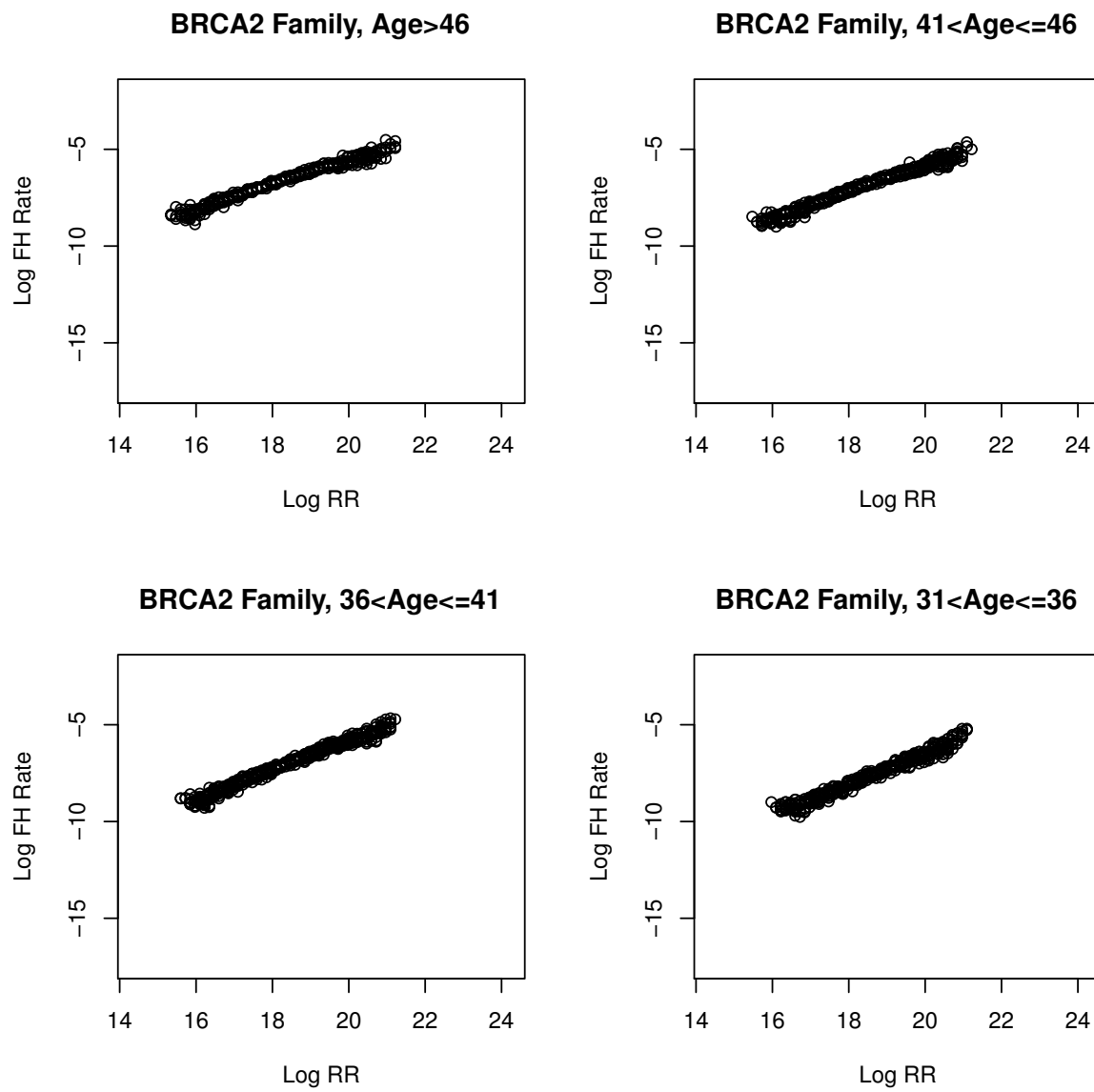


Figure 11: Logarithm of family history onset rates for families where a parent has a BRCA2 mutation. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.

To calculate benefit outgo and premium income, it is necessary to first calculate the probability of uninsured lives being in insured states at time $0 < t < 40$ (or equivalently between ages 20 and 60). We denote these probabilities, with consideration for genotype, as ${}_t^g p_x^{jk}$, *i.e.*

$${}_t^g p_x^{jk} = P(\text{In State } k \text{ at age } x+t | G^* = g \text{ and in State } j \text{ at age } x). \quad (23)$$

They may be calculated by solving the Kolmogorov forward equations:

$$\frac{d}{dt} {}_t^g p_{20}^{jk} = \sum_{l \neq k} \left({}_t^g p_{20}^{jl} \mu_{20+t}^{lk} - {}_t^g p_{20}^{jk} \mu_{20+t}^{kl} \right). \quad (24)$$

This was done using a 4th order Runge Kutta algorithm with a step size of 2^{-12} and boundary conditions: ${}_0^g p_{20}^{F1,F1} = 1$ and ${}_0^g p_{20}^{Fj,Fk} = 0$ for $(j, k) \neq (1, 1)$ and similarly for males.

We introduce notation for the critical illness hazard rate at age x , which is the same for any state: $\alpha_{g,x}^F$ for a female with genotype g and $\alpha_{g,x}^M$ for males; and $\bar{\alpha}_x^j$ as the mean critical illness hazard rate for lives in state j with weightings equal to occupancy probabilities assuming no adverse selection.

The continuously payable risk premium rate at time t may then be found using a similar technique to that used to the premiums in equation 15, by calculating the mean critical illness hazard rate for female lives in ST , $\bar{\alpha}_{x+t}^{F3}$, and including lives in FH when averaging across the sexes:

$$\Pi^{ST}(t) = \frac{\sum_{g \in \mathcal{G}^*} p_g \left[p_f \bar{\alpha}_{20+t}^{F3} \left({}_t^g p_{20}^{F3} + {}_t^g p_{20}^{F6} \right) + p_m \bar{\alpha}_{x+t}^{M3} \left({}_t^g p_{20}^{M3} + {}_t^g p_{20}^{M6} \right) \right]}{\sum_{g \in \mathcal{G}^*} p_g \left[p_f \left({}_t^g p_{20}^{F3} + {}_t^g p_{20}^{F6} \right) + p_m \left({}_t^g p_{20}^{M3} + {}_t^g p_{20}^{M6} \right) \right]}. \quad (25)$$

Lives in states $M3$ and $F3$ (see Figures 8 and 7) belong to the same underwriting class, ST , due to the requirement of unisex rates. However, the question over how to treat males with a family history among their female relatives is still present. Again there are two possible cases:

(a) Only females underwritten (lives in $M6$ are in ST underwriting class)

$$\Pi^{FH}(t) = \bar{\alpha}_{20+t}^{F6} - \bar{\alpha}_{20+t}^{F3} + \Pi^{ST}(t); \quad (26)$$

(b) Males and females are both underwritten (lives in $M6$ are in the FH underwriting class)

$$\Pi^{FH}(t) = \frac{\sum_{g \in \mathcal{G}^*} {}_t^g p_{20}^{F6} p_g p_f \left(\bar{\alpha}_{20+t}^{F6} - \bar{\alpha}_{20+t}^{F3} \right)}{\sum_{g \in \mathcal{G}^*} p_g \left({}_t^g p_{20}^{F6} p_f + {}_t^g p_{20}^{M6} p_m \right)} + \Pi^{ST}(t). \quad (27)$$

While these two possibilities may present difficulty to an insurer's underwriting department, the proportion of lives that develop a family history represents only 0.09% of male lives at its peak at age 50 from Table 15, so when testing is available to all lives, the impact of underwriting is negligible and we give results for females being underwritten.

The expected present value of benefits and premiums needed for equation 22 are then calculated with occupancy probabilities found in the presence of adverse selection, by numerical integration using Simpson's formula of the following:

$$\begin{aligned} E(\text{PV Benefit} | \text{Adverse Selection}) &= \sum_{g \in \mathcal{G}^*} p_g \left[p_f \int_0^{40} e^{-\delta t} \left({}_t^g p_{20}^{F1,ST} + {}_t^g p_{20}^{F1,FH} \right) \alpha_{g,20+t}^F dt \right. \\ &\quad \left. + p_m \int_0^{40} \left({}_t^g p_{20}^{M1,ST} + {}_t^g p_{20}^{M1,FH} \right) \alpha_{g,20+t}^M dt \right] \end{aligned} \quad (28)$$

and

$$\begin{aligned} E(\text{PV Premium} | \text{Adverse Selection}) &= \sum_{g \in \mathcal{G}^*} p_g \left[p_f \int_0^{40} e^{-\delta t} \left({}_t^g p_{20}^{F1,ST} + {}_t^g p_{20}^{M1,ST} \right) \Pi^{ST}(t) dt \right. \\ &\quad \left. + p_m \int_0^{40} e^{-\delta t} \left({}_t^g p_{20}^{F1,FH} + {}_t^g p_{20}^{M1,FH} \right) \Pi^{FH}(t) dt \right], \end{aligned} \quad (29)$$

Table 19: Cost of severe adverse selection. Testing for PG only is available to all lives, regardless of the presence of a family history of BC or OC. Underwriting is performed on females only.

Standard Insurance Rate	Test Rate	Low Risk Insurance Rate	Threshold						
			0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	High	0.050	1.430	1.326	1.042	0.691	0.385	0.179	0.069
		0.025	2.839	2.584	2.010	1.332	0.752	0.360	0.146
		0.000	6.178	5.421	4.084	2.646	1.478	0.711	0.293
	Medium	0.050	0.959	0.891	0.701	0.465	0.259	0.120	0.046
		0.025	1.855	1.693	1.321	0.878	0.497	0.238	0.096
		0.000	3.785	3.361	2.568	1.687	0.952	0.460	0.190
	Low	0.050	0.715	0.664	0.522	0.346	0.192	0.089	0.034
		0.025	1.366	1.247	0.974	0.649	0.367	0.176	0.071
		0.000	2.709	2.418	1.859	1.228	0.696	0.338	0.140
0.01	High	0.010	8.711	9.179	7.900	5.504	3.124	1.449	0.547
		0.005	11.212	11.588	9.786	6.708	3.774	1.755	0.673
		0.000	14.629	14.803	12.210	8.191	4.548	2.111	0.818
	Medium	0.010	6.365	6.555	5.528	3.802	2.143	0.990	0.372
		0.005	7.981	8.068	6.701	4.559	2.560	1.190	0.456
		0.000	10.062	9.977	8.142	5.463	3.047	1.419	0.551
	Low	0.010	4.966	5.038	4.196	2.865	1.608	0.741	0.278
		0.005	6.142	6.121	5.033	3.409	1.912	0.888	0.340
		0.000	7.608	7.450	6.038	4.049	2.263	1.056	0.410

with a step size of 2^{-11} and force of interest, $\delta = 0.05$.

5.4 Results

The results of these scenarios are shown in Tables 20 to 22. These tables show the impact of successively adding sources of adverse selection as follows:

- High risk lives adversely select at a Severe rate after a polygene test (Table 19).
- Low risk lives change behaviour after a polygene test, while high lives risk buy at the standard rate (Table 20).
- High risk lives adversely select at a Severe rate after tests for polygene and BRCA — lives with a low risk polygene but with BRCA1 or BRCA2 mutations also adversely select (Table 21).
- High risk lives adversely select at a Moderate rate after tests for polygene and BRCA — lives with a low risk polygene but with BRCA1 or BRCA2 mutations also adversely select (Table 22).
- High risk lives adversely select at a Severe rate after tests for polygene and BRCA — lives with a low risk polygene but with BRCA1 or BRCA2 mutations also adversely select. Tests available only to lives with a family history of breast or ovarian cancer (Table 23).

The costs involved look comparable to those in Macdonald & McIvor (2009). However our model differs from theirs in a number of ways, key of which is the European Union ruling described in Section 4. This spreads the costs over approximately twice the volume of premium. Thus the costs suggested by our model would need to be approximately doubled to apply to a regime where sexual discrimination is permitted. Where this is the case, our somewhat moderate costs could become quite substantial particularly when passed on to the customer.

One observation of note is that a large part of the cost of adverse selection could come from low risk lives buying at a lower rate after testing. Although the reduction in the portfolio of insured lives by any individual low risk life increases the average risk by a smaller amount than the gain of a high risk

Table 20: Cost of adverse selection when high-risk lives buy at the standard rate. Testing for PG only is available to all lives, regardless of the presence of a family history of BC or OC. Underwriting is performed on females only.

Standard Insurance Rate	Test Rate	Low Risk Insurance Rate	Threshold						
			0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	High	0.025	1.361	1.212	0.937	0.627	0.362	0.180	0.077
		0.000	4.623	3.960	2.951	1.913	1.079	0.529	0.223
	Medium	0.025	0.875	0.782	0.607	0.408	0.236	0.117	0.050
		0.000	2.770	2.412	1.829	1.204	0.687	0.339	0.144
	Low	0.025	0.639	0.572	0.445	0.299	0.174	0.087	0.037
		0.000	1.964	1.723	1.317	0.873	0.501	0.248	0.105
0.01	High	0.005	2.374	2.066	1.569	1.035	0.592	0.292	0.124
		0.000	5.836	4.915	3.611	2.316	1.297	0.632	0.267
	Medium	0.005	1.527	1.338	1.024	0.680	0.391	0.194	0.082
		0.000	3.552	3.052	2.289	1.494	0.847	0.416	0.176
	Low	0.005	1.113	0.979	0.752	0.501	0.289	0.143	0.061
		0.000	2.526	2.190	1.658	1.092	0.623	0.307	0.130

Table 21: Cost of severe adverse selection. Testing for MG and PG is available to all lives, regardless of presence of a family history of BC or OC. Underwriting is performed on females only.

Standard Insurance Rate	Test Rate	Low Risk Insurance Rate	Threshold						
			0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	High	0.050	1.462	1.373	1.103	0.763	0.464	0.262	0.153
		0.025	2.893	2.646	2.080	1.410	0.833	0.444	0.230
		0.000	6.278	5.512	4.172	2.733	1.565	0.797	0.378
	Medium	0.050	0.982	0.923	0.742	0.513	0.311	0.175	0.102
		0.025	1.891	1.734	1.367	0.930	0.551	0.294	0.152
		0.000	3.846	3.418	2.625	1.744	1.009	0.517	0.247
	Low	0.050	0.732	0.688	0.552	0.382	0.232	0.130	0.076
		0.025	1.392	1.277	1.008	0.687	0.407	0.217	0.113
		0.000	2.753	2.459	1.901	1.270	0.738	0.380	0.182
0.01	High	0.010	8.913	9.490	8.337	6.045	3.729	2.085	1.196
		0.005	11.448	11.932	10.252	7.271	4.392	2.398	1.325
		0.000	14.909	15.189	12.712	8.780	5.182	2.762	1.473
	Medium	0.010	6.516	6.782	5.837	4.176	2.557	1.423	0.813
		0.005	8.155	8.314	7.026	4.945	2.981	1.626	0.897
		0.000	10.263	10.247	8.486	5.862	3.475	1.860	0.994
	Low	0.010	5.086	5.214	4.432	3.147	1.919	1.066	0.608
		0.005	6.278	6.310	5.279	3.698	2.227	1.214	0.670
		0.000	7.763	7.655	6.296	4.346	2.582	1.384	0.741

Table 22: Cost of moderate adverse selection. Testing for MG and PG is available to all lives, regardless of presence of family history of BC or OC. Underwriting is performed on females only.

Standard Insurance Rate	Test Rate	Low Risk Insurance Rate	Threshold						
			0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	High	0.050	0.818	0.781	0.632	0.438	0.265	0.148	0.085
		0.025	2.225	2.033	1.596	1.078	0.632	0.329	0.162
		0.000	5.569	4.857	3.660	2.389	1.359	0.681	0.310
	Medium	0.050	0.536	0.513	0.416	0.288	0.174	0.097	0.056
		0.025	1.435	1.314	1.035	0.702	0.413	0.215	0.106
		0.000	3.372	2.981	2.281	1.510	0.869	0.438	0.200
	Low	0.050	0.396	0.379	0.307	0.213	0.129	0.072	0.041
		0.025	1.050	0.963	0.760	0.516	0.304	0.159	0.078
		0.000	2.401	2.136	1.646	1.097	0.634	0.321	0.147
0.01	High	0.010	7.140	7.609	6.651	4.795	2.939	1.629	0.925
		0.005	9.643	9.984	8.501	5.983	3.588	1.938	1.053
		0.000	13.092	13.166	10.881	7.447	4.360	2.297	1.200
	Medium	0.010	5.085	5.312	4.563	3.252	1.980	1.093	0.618
		0.005	6.698	6.805	5.718	4.002	2.397	1.294	0.702
		0.000	8.783	8.692	7.136	4.897	2.883	1.525	0.798
	Low	0.010	3.918	4.041	3.435	2.433	1.476	0.813	0.459
		0.005	5.090	5.112	4.262	2.973	1.780	0.961	0.521
		0.000	6.556	6.427	5.255	3.609	2.130	1.129	0.592

life, the volume of low risk lives underbuying compared to high risk lives overbuying causes this high cost relating to low risk lives.

Adding BRCA to the test regime makes only a small impact when already testing for polygenes. This validates the overall finding of Macdonald & McIvor (2009), that the bigger part of adverse selection cost is down to the polygene.

As previously stated, very few lives develop a family history. If testing is limited only to those who have a family history then the cost from adverse selection is negligible, reaching a peak of 0.003% in a small market with High test rate.

As the science develops and more genes are identified, the feasibility and cost of testing will improve. Pharoah *et al.* (2008) discusses the use of population testing as a way of identifying high-risk women in order to target the more expensive means of detecting breast cancers, *e.g.* MRI screening, for early diagnosis. If the testing is extended in this way the potential for anti-selection increases greatly, with most lives being aware of their risk profile. Early detection will however change the shape of rate tables and these early cases could be at a stage that is not covered under the critical illness cover *e.g.* ductal carcinoma in situ is not considered a critical illness by many insurers, but left untreated may develop into invasive breast cancer. Where the prognosis improves such that a claim would not be viable, the inclination to adversely select would decrease so we consider that the Moderate adverse selection rate to reflect this case.

In a small market, under the Moderate adverse selection scenarios there are still fairly moderate costs involved, up to 14%. Indeed the change from the Severe rate, is quite small as compared to the standard rate of insurance, this is already a severe rate of uptake.

6. CONCLUSIONS

The discovery of a small number of ‘breast cancer polygenes’ has allowed us to take another look at Macdonald & McIvor (2006)’s work which previously used a hypothetical model in order to assess the

Table 23: Cost of severe adverse selection. Testing for MG and PG is available only after the development of a family history of BC or OC. Underwriting is performed on females only.

Standard Insurance Rate	Test Rate	Low Risk Insurance Rate	Threshold						
			0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	High	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Medium	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Low	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.01	High	0.010	0.001	0.002	0.002	0.003	0.002	0.002	0.002
		0.005	0.002	0.002	0.002	0.003	0.002	0.002	0.002
		0.000	0.002	0.002	0.002	0.003	0.002	0.002	0.002
	Medium	0.010	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Low	0.010	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001

findings and update their results as necessary. They had suggested the simplifications of that model create a heavy tail that distorts the costs. We have found that their results for family history ratings, when corrected to use the standard premium as numeraire, are reasonable. Additionally their conclusion that the bigger part of adverse selection risk is attributable to polygenes still stands in a model based on actual genetic data. However they understate the size of the adverse selection cost compared to that arising under our actual gene model.

The change of European law to enforce the use of unisex pricing, while creating additional concerns for an insurer's underwriting team and adding a level of risk from business mix, acts to spread the cost of adverse selection over a wider base whose behaviour is not subject to change. The premiums chargeable to a female with family history would be roughly halved if the same rating was applied to a male despite the lack of additional risk implied for the male.

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 the Continuous Mortality Investigation for providing funding.

We would also like to thank Dr. Kenneth McIvor for his help in reconciling our respective models.

REFERENCES

- ANTONIOU, A., PHAROAH, P., McMULLAN, G., DAY, N., PONDER, B., & EASTON, D. (2001). Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology*, **21**, pp. 1–18.
- ANTONIOU, A., PHAROAH, P., McMULLAN, G., DAY, N., STRATTON, M., PETO, J., PONDER, B., & EASTON, D. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer*, **86**, pp. 76–83.

- ANTONIOU, A.C., SPURDLE, A.B., SINILNIKOVA, O.M., HEALEY, S., POOLEY, K.A., SCHMUTZLER, R.K., VERSMOLD, B., ENGEL, C., MEINDL, A., ARNOLD, N., HOFMANN, W., SUTTER, C., NIEDERACHER, D., DESSLER, H., CALDES, T., KAMPJARVI, K., NEVANLINNA, H., SIMARD, J., BEESLEY, J., CHEN, X., NEUHAUSEN, S.L., REBBECK, T.R., WAGNER, T., LYNCH, H.T., ISAACS, C., WEITZEL, J., GANZ, P.A., DALY, M.B., TOMLINSON, G., OLOPADE, O.I., BIUM, J.L., COUCH, F.J., PETERLONGO, P., MANOUKIAN, S., BARILE, M., RADICE, P., SZABO, C.I., PEREIRA, L.H.M., GREENE, M.H., RENNERT, G., LEIBKOWICZ, F., BARNETT-GRINNESS, O., ANDRULIS, I.L., OZCELIK, H., GERDES, A.M., CALIGO, M.A., LAITMAN, Y., KAUFMAN, B., MILGROM, R., FRIEDMAN, E., DOMCHEK, S.M., NATHANSON, K.L., OSORIO, A., LLORT, G., MILNE, R.L., BENITEZ, J., HAMANN, U., HOGERVORST, F.B.L., MANDERS, P., LIGTENBERG, M.J.L., VAN DEN OUWELAND, A.M.W., PEOCK, S., COOK, M., PLATTE, R., EVANS, D.G., EELES, R., PICHERT, G., CHU, C., ECCLES, D., DAVIDSON, R., DOUGLAS, F., GODWIN, A.K., BARJHOUS, L., MAZOYER, S., SOBOL, H., BOURDON, V., EISINGER, F., CHOMPRET, A., CAPOULADE, C., PAILLERETS, B.B.D., LENOIR, G.M., GAUTHIER-VILLARS, M., HOUDAYER, C., STOPPA-LYONNET, D., EASTON, D.F., RES, K.C.C., OCGN, SWEDISH BRCA1 AND DNA-HEBON COLLABORATORS, EMBRACE, GEMO, & CIMBA (2008). Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *American Journal of Human Genetics*, **82**, pp. 937–948.
- BROWN, M. (1992). Evidence of adverse selection in the individual health insurance market. *The Journal of Risk and Insurance*, **59**, pp. 13–33.
- COX, A., DUNNING, A.M., GARCIA-CLOSAS, M., BALASUBRAMANIAN, S., REED, M.W., POOLEY, K.A., SCOLLEN, S., BAYNES, C., PONDER, B.A., CHANOCK, S., LISSOWSKA, J., BRINTON, L., PEPLONSKA, B., SOUTHEY, M.C., HOPPER, J.L., MCCREDIE, M.R., GILES, G.G., FLETCHER, O., JOHNSON, N., DOS SANTOS SILVA, I., GIBSON, L., BOJESSEN, S.E., NORDESTGAARD, B.G., AXELSSON, C.K., TORRES, D., HAMANN, U., JUSTENHOVEN, C., BRAUCH, H., CHANG-CLAUDE, J., KROPP, S., RISCH, A., WANG-GOHRKE, S., SCHUERMANN, P., BOGDANOVA, N., DOERK, T., FAGERHOLM, R., AALTONEN, K., BLOMQUIST, C., NEVANLINNA, H., SEAL, S., RENWICK, A., STRATTON, M.R., RAHMAN, N., SANGRAJRANG, S., HUGHES, D., ODEFREY, F., BRENNAN, P., SPURDLE, A.B., CHENEVIX-TRENCH, G., BEESLEY, J., MANNERMAA, A., HARTIKAINEN, J., KATAJA, V., KOSMA, V.M., COUCH, F.J., OLSON, J.E., GOODE, E.L., BROEKS, A., SCHMIDT, M.K., HOGERVORST, F.B.L., VAN'T VEER, L.J., KANG, D., YOO, K.Y., NOH, D.Y., AHN, S.H., WEDREN, S., HALL, P., LOW, Y.L., LIU, J., MILNE, R.L., RIBAS, G., GONZALEZ-NEIRA, A., BENITEZ, J., SIGURDSON, A.J., STREDRICK, D.L., ALEXANDER, B.H., STRUEWING, J.P., PHAROAH, P.D., EASTON, D.F., KATHLEEN CUNNINGHAM FDN CONSORTIUM, & BREAST CANC ASSOC CONSORTIUM (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*, **39:3**, pp. 352–358.
- EASTON, D.F., POOLEY, K.A., DUNNING, A.M., PHAROAH, P.D.P., THOMPSON, D., BALLINGER, D.G., STRUEWING, J.P., MORRISON, J., FIELD, H., LUBEN, R., WAREHAM, N., AHMED, S., HEALEY, C.S., BOWMAN, R., MEYER, K.B., HAIMAN, C.A., KOLONEL, L.K., HENDERSON, B.E., LE MARCHAND, L., BRENNAN, P., SANGRAJRANG, S., GABORIEAU, V., ODEFREY, F., SHEN, C.Y., WU, P.E., WANG, H.C., ECCLES, D., EVANS, D.G., PETO, J., FLETCHER, O., JOHNSON, N., SEAL, S., STRATTON, M.R., RAHMAN, N., CHENEVIX-TRENCH, G., BOJESSEN, S.E., NORDESTGAARD, B.G., AXELSSON, C.K., GARCIA-CLOSAS, M., BRINTON, L., CHANOCK, S., LISSOWSKA, J., PEPLONSKA, B., NEVANLINNA, H., FAGERHOLM, R., EEROLA, H., KANG, D., YOO, K.Y., NOH, D.Y., AHN, S.H., HUNTER, D.J., HANKINSON, S.E., COX, D.G., HALL, P., WEDREN, S., LIU, J., LOW, Y.L., BOGDANOVA, N., SCHUERMANN, P., DOERK, T., TOLLENAAR, R.A.E.M., JACOBI, C.E., DEVILEE, P., KLIJN, J.G.M., SIGURDSON, A.J., DOODY, M.M., ALEXANDER, B.H., ZHANG, J., COX, A., BROCK, I.W., MACPHERSON, G., REED, M.W.R., COUCH, F.J., GOODE, E.L., OLSON, J.E., MEIJERS-HEIJBOER, H., VAN DEN OUWELAND, A., UITTERLINDEN, A., RIVADENEIRA, F., MILNE, R.L., RIBAS, G., GONZALEZ-NEIRA, A., BENITEZ, J., HOPPER, J.L., MCCREDIE, M., SOUTHEY, M., GILES, G.G., SCHROEN, C., JUSTENHOVEN, C., BRAUCH, H., HAMANN, U., KO, Y.D., SPURDLE, A.B., BEESLEY, J., CHEN, X., MANNERMAA, A., KOSMA, V.M., KATAJA, V., HARTIKAINEN, J., DAY, N.E., COX, D.R., PONDER, B.A.J., SEARCH COLLABORATORS, kCONFAB, & AOCS MANAGEMENT GRP (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, pp. 1087–1095.
- EUROPEAN COURT OF JUSTICE (2011). C-236/09 Association Belge des Consommateurs Test-Achats ASBL. **ECR I-00773**.
- FINKELSTEIN, A. & MCGARRY, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economics Review*, **96**, pp. 938–958.
- FINKELSTEIN, A. & POTERBA, J. (2002). Selection effects in the United Kingdom individual annuities market. *The Economic Journal*, **112:476**, pp. 28–50.

- GHOUSSAINI, M., FLETCHER, O., MICHAILEDIOU, K., TURNBULL, C., SCHMIDT, M.K., DICKS, E., DENNIS, J., WANG, Q., HUMPHREYS, M.K., LUCCARINI, C., BAYNES, C., CONROY, D., MARANIAN, M., AHMED, S., DRIVER, K., JOHNSON, N., ORR, N., DOS SANTOS SILVA, I., WAISFISZ, Q., MEIJERS-HEIJBOER, H., UITTERLINDEN, A.G., RIVADENEIRA, F., HALL, P., CZENE, K., IRWANTO, A., LIU, J., NEVANLINNA, H., AITTO MAKI, K., BLOMQUIST, C., MEINDL, A., SCHMUTZLER, R.K., MULLER-MYHSOK, B., LICHTNER, P., CHANG-CLAUDE, J., HEIN, R., NICKELS, S., FLESCH-JANYS, D., TSIMIKLIS, H., MAKALIC, E., SCHMIDT, D., BUI, M., HOPPER, J.L., APICELLA, C., PARK, D.J., SOUTHEY, M., HUNTER, D.J., CHANOCK, S.J., BROEKS, A., VERHOEF, S., HOGERVORST, F.B., FASCHING, P.A., LUX, M.P., BECKMANN, M.W., EKICI, A.B., SAWYER, E., TOMLINSON, I., KERIN, M., MARME, F., SCHNEEWEISS, A., SOHN, C., BURWINKEL, B., GUENEL, P., TRUONG, T., CORDINA-DUVERGER, E., MENEGAUX, F., BOJESSEN, S.E., NORDESTGAARD, B.G., NIELSEN, S.F., FLYGER, H., MILNE, R.L., ALONSO, M.R., GONZALEZ-NEIRA, A., BENITEZ, J., ANTON-CULVER, H., ZIOGAS, A., BERNSTEIN, L., DUR, C.C., BRENNER, H., MULLER, H., ARNDT, V., STEGMAIER, C., JUSTENHOVEN, C., BRAUCH, H., BRUNING, T., WANG-GOHRKE, S., EILBER, U., DORK, T., SCHURMANN, P., BREMER, M., HILLEMANN, P., BOGDANOVA, N.V., ANTONENKOVA, N.N., ROGOV, Y.I., KARSTENS, J.H., BERMISHEVA, M., PROKOFIEVA, D., KHUSNUTDINOVA, E., LINDBLOM, A., MARGOLIN, S., MANNERMAA, A., KATAJA, V., KOSMA, V.M., HARTIKAINEN, J.M., LAMBRECHTS, D., YESILYURT, B.T., FLORIS, G., LEUNEN, K., MANOUKIAN, S., BONANNI, B., FORTUZZI, S., PETERLONGO, P., COUCH, F.J., WANG, X., STEVENS, K., LEE, A., GILES, G.G., BAGLIETTO, L., SEVERI, G., MCLEAN, C., ALNAES, G.G., KRISTENSEN, V., BORRENSEN-DALE, A.L., JOHN, E.M., MIRON, A., WINQVIST, R., PYLKAS, K., JUUKOLA-VUORINEN, A., KAUPPILA, S., ANDRULIS, I.L., GLENDON, G., MULLIGAN, A.M., DEVILEE, P., VAN ASPEREN, C.J., TOLLENAAR, R.A., SEYNAEVE, C., FIGUEROA, J.D., GARCIA-CLOSAS, M., BRINTON, L., LISSOWSKA, J., HOONING, M.J., HOLLESTELLE, A., OLDENBURG, R.A., VAN DEN OUWELAND, A.M., COX, A., REED, M.W., SHAH, M., JAKUBOWSKA, A., LUBINSKI, J., JAWORSKA, K., DURDA, K., JONES, M., SCHOEMAKER, M., ASHWORTH, A., SWERDLOW, A., BEESLEY, J., CHEN, X., MUIR, K.R., LOPHATANANON, A., RATTANAMONGKONGUL, S., CHAIWERAWATTANA, A., KANG, D., YOO, K.Y., NOH, D.Y., SHEN, C.Y., YU, J.C., WU, P.E., HSIUNG, C.N., PERKINS, A., SWANN, R., VELENTZIS, L., ECCLES, D.M., TAPPER, W.J., GERTY, S.M., GRAHAM, N.J., PONDER, B.A., CHENEVIX-TRENCH, G., PHAROAH, P.D., LATHROP, M., DUNNING, A.M., RAHMAN, N., PETO, J., & EASTON, D. (2012). Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nature Genetics*, **44**:3, pp. 312–318.
- 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 with insurance applications. *Scandinavian Actuarial Journal*, **6**, pp. 413–419.
- GUTIÉRREZ, C. & MACDONALD, A.S. (2003). Adult polycystic kidney disease and critical illness insurance. *North American Actuarial Journal*, **7**:2, pp. 93–115.
- HUNTER, D.J., KRAFT, P., JACOBS, K.B., COX, D.G., YEAGER, M., HANKINSON, S.E., WACHOLDER, S., WANG, Z., WELCH, R., HUTCHINSON, A., WANG, J., YU, K., CHATTERJEE, N., ORR, N., WILLETT, W.C., COLDITZ, G.A., ZIEGLER, R.G., BERG, C.D., BUYS, S.S., MCCARTY, C.A., FEIGELSON, H.S., CALLE, E.E., THUN, M.J., HAYES, R.B., TUCKER, M., GERHARD, D.S., FRAUMENI, JR., J.F., HOOVER, R.N., THOMAS, G., & CHANOCK, S.J. (2007). A genome-wide association study identifies alleles in *fgfr2* associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics*, **39**, pp. 870–874.
- LANGE, K. (1997). An approximate model of polygenic inheritance. *Genetics*, **147**, pp. 1423–1430.
- LEMAIRE, J., SUBRAMANIAN, K., ARMSTRONG, K., & ASCH, D. (2000). Pricing term insurance in the presence of a family history of breast or ovarian cancer. *North American Actuarial Journal*, **4**, pp. 75–87.
- LU, B., MACDONALD, A.S., & WATERS, H.R. (2011). The genetics of breast and ovarian cancer IV: A model of breast cancer progression. *Scandinavian Actuarial Journal*, **6**, pp. 239–266.
- LU, B., MACDONALD, A.S., WATERS, H.R., & YU, F. (2011). The genetics of breast and ovarian cancer V: Application to income protection insurance. *Scandinavian Actuarial Journal*, **4**, pp. 267–291.
- MACDONALD, A.S. & MCIIVOR, K. (2006). Application of a polygenic model of breast and ovarian cancer to critical illness insurance. *Annals of Actuarial Science*, **1**:2, pp. 319–343.
- MACDONALD, A.S. & MCIIVOR, K. (2009). Modelling adverse selection in the presence of a common genetic disorder: The breast cancer polygene. *ASTIN Bulletin*, **39**, pp. 373–402.

- 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*, **2003**, pp. 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*, **2003**, pp. 28–50.
- NAROD, S., NEUHAUSEN, S., VICHODEZ, G., ARMEL, S., LYNCH, H., GHADIRIAN, P., CUMMINGS, S., OLOPADE, O., STOPPA-LYONNET, D., COUCH, F., WAGNER, T., WARNER, E., FOULKES, W., SAAL, H., WEITZEL, J., TULMAN, A., POLL, A., NAM, R., SUN, P., & HEREDITARY BREAST CANCER STUDY GRP (2008). Rapid progression of prostate cancer in men with a BRCA2 mutation. *British Journal of Cancer*, **99:2**, pp. 371–374.
- PARKIN, D., BRAY, F., & DEVESEA, S. (2001). Cancer burden in the year 2000. the global picture. *European Journal of Cancer*, **37, Supplement 8**, pp. 4–66.
- PHAROAH, P.D., ANTONIOU, A.C., EASTON, D.F., & PONDER, B.A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *New England Journal of Medicine*, **358**, pp. 2796–2803.
- ROPKA, M., WENZEL, J., PHILLIPS, E., SIADATY, M., & PHILBRICK, J. (2006). Uptake rates for breast cancer genetic testing: A systematic review. *Cancer Epidemiology, Biomarkers, and Prevention*, **15:5**, pp. 840–855.
- STACEY, S.N., MANOLESCU, A., SULEM, P., RAFNAR, T., GUDMUNDSSON, J., GUDJONSSON, S.A., MASSON, G., JAKOBSDOTTIR, M., THORLACIUS, S., HELGASON, A., ABEN, K.K., STROBBE, L.J., ALBERS-AKKERS, M.T., SWINKELS, D.W., HENDERSON, B.E., KOLONEL, L.N., LE MARCHAND, L., MILLASTRE, E., ANDRES, R., GODINO, J., GARCIA-PRATS, M.D., POLO, E., TRES, A., MOUY, M., SAEMUNSDOTTIR, J., BACKMAN, V.M., GUDMUNDSSON, L., KRISTJANSSON, K., BERGTHORSSON, J.T., KOSTIC, J., FRIGGE, M.L., GELLER, F., GUDBJARTSSON, D., SIGURDSSON, H., JONSDOTTIR, T., HRAFNKELSSON, J., JOHANNSSON, J., SVEINSSON, T., MYRDAL, G., GRIMSSON, H.N., JONSSON, T., VON HOLST, S., WERELIUS, B., MARGOLIN, S., LINDBLOM, A., MAYORDOMO, J.I., HAIMAN, C.A., KIEMENEY, L.A., JOHANNSSON, O.T., GULCHER, J.R., THORSTEINSDOTTIR, U., KONG, A., & STEFANSSON, K. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*, **39:7**, pp. 865–869.
- SUBRAMANIAN, K., LEMAIRE, J., HERSHEY, J., PAULY, M., ARMSTRONG, K., & ASCH, D. (1999). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *Journal of Risk and Insurance*, **66:4**, pp. 531–550.
- TAI, Y.C., DOMCHEK, S., PARMIGIANI, G., & CHEN, S. (2007). Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *Journal of the National Cancer Institute*, **99:23**, pp. 1811–1814.
- TUTT, A. & ASHWORTH, A. (2002). The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. *Trends in Molecular Medicine*, **8:12**, pp. 571–579.
- WASHBROOK, E. (2006). Risk factors and epidemiology of breast cancer. *Womens Health Medicine*, **3:1**, pp. 8–14.
- ZICK, C.D., SMITH, K.R., MAYER, R.N., & BOTKIN, J.R. (2000). Genetic testing, adverse selection, and the demand for life insurance. *American Journal of Medical Genetics*, **93:1**, pp. 29–39.