MULTIFACTORIAL GENETIC DISORDERS AND ADVERSE SELECTION: EPIDEMIOLOGY MEETS ECONOMICS

By Angus Macdonald and Pradip Tapadar

ABSTRACT

Rapid advances in genetic epidemiology and the setting up of large-scale cohort studies have shifted the focus from severe, but rare, single gene disorders to less severe, but common, multifactorial disorders. This will lead to the discovery of genetic risk factors for common diseases of major importance in insurance underwriting. If genetic information continues to be treated as private, adverse selection becomes possible, but it should occur only if the individuals at lowest risk obtain lower expected utility by purchasing insurance at the average price than by not insuring. We explore where this boundary may lie, using a simple 2×2 gene-environment interaction model of epidemiological risk, in a simplified 2-state insurance model and in a more realistic model of heart-attack risk and critical illness insurance. Adverse selection does not appear unless purchasers are not very risk-averse, and insure a small proportion of their wealth; or unless the elevated risks implied by genetic information are implausibly high. In many cases adverse selection is impossible if the low-risk stratum of the population is large enough. These observations are strongly accentuated in the critical illness model by the presence of risks other than heart attack, and the constraint that differential heart-attack risks must agree with the overall population risk. We find no convincing evidence that adverse selection is a serious insurance risk, even if information about multifactorial genetic disorders remains private.

CONTACT ADDRESS

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

1. Introduction

1.1 Risk and Insurance

The principle behind underwriting is to identify key risk factors that stratify applicants into reasonably homogeneous groups, for each of which the appropriate premium rate can be charged. The risk of death or ill health is affected by, among other things, age, gender, lifestyle and genotype. However, the use of certain risk factors is sometimes controversial. In particular, this is true of factors over which individuals have no control, such as genotype. As a result, in many countries a ban has been imposed, or moratorium agreed, limiting the use of genetic information. In one country, the UK, a government-appointed Genetics and Insurance Committee (GAIC) is providing guidance to insurers on the acceptable use of genetic test results.

Disorders caused by mutations in single genes, which may be severe and of late onset, but are rare, have been quite extensively studied in the insurance literature, see Macdonald (2004) for a review. One reason is that the epidemiology of these disorders is relatively advanced, because biological cause and effect could be traced relatively easily. The conclusion has been that single-gene disorders, because of their rarity, do not expose insurers to serious adverse selection in large enough markets.

The vast majority of the genetic contribution to human disease, however, will arise from combinations of gene varieties (called 'alleles') and environmental factors, each of which might be quite common, and each alone of small influence but together exerting a measurable effect on the molecular mechanism of a disease. Some combinations may be protective, others deleterious. These are the multifactorial disorders, and they are the future of genetics research. Their epidemiology is not very advanced, but should make progress in the next 5–10 years through the very large prospective studies now beginning in several countries. One of the largest is the Biobank project in the UK, with 500,000 subjects, described in Macdonald, Pritchard & Tapadar (2006). UK Biobank will recruit 500,000 people aged 40 to 69 from the general population of the UK, and follow them up for 10 years. The aim is to capture both genetic and environmental variations and interactions, and relate them to the risks of common diseases. If successful, the outcome will be much better knowledge of the risks associated with complex genotypes. Thus the genetics and insurance debate will, in the fairly near future, shift from single-gene to multifactorial disorders.

Any model used to study adverse selection risk must incorporate the behaviour of the market participants. Most of those applied to single-gene disorders in the past did so in a very simple and exaggerated way, assuming that the risk implied by an adverse genetic test result was so great that its recipient would quickly buy life or health insurance with very high probability. These assumptions were not based on any quantified economic rationale, but since they led to minimal changes in the price of insurance this probably did not matter. The same is not true if we try to model multifactorial disorders. Then 'adverse' genotypes may imply relatively modest excess risk but may be reasonably common, so the decision to buy insurance is more central to the outcome.

Subramanian et al. (1999) used a continuous-time discrete-state Markov model to estimate adverse selection costs for term insurance contracts resulting from non-disclosure of BRCA mutation test results and/or a family history of breast and ovarian cancer. This was the first study explicitly to link adverse selection and genetic epidemiology. They assumed that cover would be increased if a genetic test reveals higher risk, and reduced if it does not. The cost of adverse selection was defined as the across-the-board increase in premiums needed for an insurer who did not observe the genetic test results to absorb the extra cost. These increases could reach 120% in scenarios where women disclosed family histories but not test results, However, they could exceed 200%, approaching 600% in extreme scenarios, when family histories were not disclosed either. The authors concluded that if companies do not identify applicants' family histories, adverse selection costs could become unbearable.

Information asymmetry and adverse selection have also been considered in an equilibrium setting. Doherty & Thistle (1996) pointed out that, under symmetric information, insurance deters diagnostic testing. This is because the premium is a lottery whose value

is revealed by the test, and risk-averse individuals will prefer a pooled premium. Asymmetric information alters or abolishes the equilibrium, depending on the cost of being tested, and whether or not low-risk individuals may choose to reveal beneficial test results. Hoy & Polborn (2000), following also Villeneuve (1996), analysed the same problem in a life insurance model, in which lost income due to premature death is replaced. They constructed scenarios where a new test could increase or decrease the social value of information.

Hoy & Witt (2007) applied the results from Hoy & Polborn (2000) to the specific case of the BRCA1/2 breast cancer genes. They simulated the market for 10-year term life insurance policies targeted at women aged 35 to 39. They stratified the consumer base into 13 risk categories based on family background information. This information is also available to insurers. Then within each risk group, they checked the impact of test results for BRCA1/2 genes on welfare effects, using iso-elastic utility functions. The authors showed that in the presence of a high risk group, *and* in the presence of information asymmetry, the equilibrium insurance premium can be as high as 297% of the population weighted probability of death, but this was very much a worst-case scenario.

Polborn, Hoy & Sadanand (2006) developed a model where individuals, early in their lives, know neither the levels of insurance they will demand later in life, nor their mortality risk, which they learn over time. Under this set-up, the characteristics of the equilibrium level of initial insurance purchase are derived, assuming both symmetric and asymmetric information. The authors show that, under certain assumptions, regulations prohibiting the use of genetic test information will increase welfare despite creating adverse selection. This implies that individuals would prefer to face adverse selection costs rather than premium risks.

Hoy (2006) concentrates on the social welfare issues related to risk classification. In particular, he asks whether regulations that create adverse selection improve or worsen expected welfare. Social welfare is affected by adverse selection costs on one hand and protection against premium risk on the other. The author concludes that, on balance, if the proportion of high-risk types within the population exceeds a certain threshold, then regulatory adverse selection unambiguously reduces expected welfare. However, if the proportion of high-risk individuals is sufficiently small, then welfare can be enhanced by banning risk classification. Although we do not address social welfare issues in this paper, we will obtain the threshold proportion of high-risk types, above which the pooled insurance premium will become unacceptably high for low-risk individuals.

All these papers assume that the genetic epidemiology implies that genetic tests carry very strong information about risk; true of some single-gene disorders but unlikely to be so true of multifactorial disorders. They concentrate primarily on providing a proper economic rationale for the impact, on the insurance market, of genetic tests for, mainly, rare diseases. In this paper, we try to bring together plausible quantitative models for the epidemiology and the economic issues, in respect of more common disorders, therefore affecting a much larger proportion of the insurer's customer base. We wish to find out under what circumstances adverse selection is likely to occur with sufficient force to be problematic.

We suppose that individuals are risk-averse, have wealth W and aim to buy insurance with sum assured $L \leq W$. Their decision is governed by expected utility, conditioned on

the information available to them. Insurers, in a competitive market, charge an actuarially fair premium P, equal to the expected present value of the insured loss, conditioned on the information available to them. See for example Hoy & Polborn (2000) for a similar market model. Because they are risk-averse, individuals will be willing to pay a premium up to a maximum of $P^* > P$, provided that they and the insurer have the same information. We can then consider the effect of genetic information that is only available to applicants.

We propose a simple model of a multifactorial disorder, with two genotypes and two levels of environmental exposure, and either additive or multiplicative interactions between them. These factors affect the risk of myocardial infarction (heart attack), therefore the theoretical price of critical illness (CI) insurance. However these price differences are not very large. To begin with, the risk factors are not observable, because the epidemiology is unknown, or the necessary genetic tests have not yet been developed. Insurers therefore charge everyone the same premium, which is the appropriate weighted average of the genotype and environment-specific premiums. Subsequently, genetic tests that accurately predict the risk become available, but only to individuals; insurers are barred from asking about genotype. Adverse selection therefore becomes a possibility.

2. UTILITY FUNCTIONS

2.1 *Utility of Wealth*

We assume that all individuals who may buy insurance have the same utility function, namely an increasing concave function $U(w)$ of wealth w (so $U'(w) > 0$ and $U''(w) < 0$). Current wealth, which is deterministic, is compared with wealth after the outcome of a probabilistic experiment *via* the expected utility of the outcome. Since the nature of the probabilistic experiment underlying insurance involves the timing as well as the occurrence of the insured event, we will measure wealth in present value terms when necessary. For a full exposition of utility theory, see Binmore (1991).

Suppose the individual with utility function $U(w)$ has initial wealth W but with probability q will lose L. Their ultimate wealth is the random variable X, where $X = W - L$ with probability q and $X = W$ with probability $1 - q$. If they choose, they can insure the risk for premium P, and accept $W - P$ with certainty. They should do so if:

$$U(W - P) > \mathrm{E}[U(X)] = qU(W - L) + (1 - q)U(W). \tag{1}$$

In particular they should insure if the premium is equal to the expected loss qL since for a risk-averse individual:

$$U(W - qL) = U(q(W - L) + (1 - q)W) > qU(W - L) + (1 - q)U(W). \tag{2}$$

So in a market where competition drives insurers to charge the actuarially 'fair' premium qL, insurance will be bought, but this is not the limiting case; insurance will be bought as long as the premium is less than P^* where:

$$P^* = W - U^{-1}[qU(W - L) + (1 - q)U(W)]. \tag{3}$$

Hence, P^* is the maximum willingness-to-pay, and $P^* - qL$ is the risk premium.

2.2 *Coefficients of Risk-aversion*

The Arrow-Pratt measure of (absolute) risk-aversion of a utility function $U(w)$ is defined as:

$$A_U(w) = -\frac{U''(w)}{U'(w)}. \tag{4}$$

It is well-known that two utility functions represent the same preference relation if and only if they have the same absolute risk-aversion function. A related quantity is the measure of relative risk-aversion, defined as:

$$R(w) = A_U(w)w = -\frac{U''(w)w}{U'(w)}. \tag{5}$$

2.3 *Families of Utility Functions*

We introduce two families of utility functions which we will use in examples through-out the rest of the paper.

(a) The Iso-Elastic utility functions are defined by:

$$U_{I(\lambda)}(w) = \begin{cases} (w^\lambda - 1)/\lambda & \lambda < 1 \text{ and } \lambda \neq 0 \\ \log(w) & \lambda = 0. \end{cases} \tag{6}$$

The condition $\lambda < 1$ ensures concavity. Log-utility is the limiting case as $\lambda \to 0$. The absolute risk-aversion function of $U_{I(\lambda)}(w)$ is:

$$A(w) = \frac{1 - \lambda}{w} \tag{7}$$

and the relative risk-aversion function is constant, $R(w) = R = 1 - \lambda$. Hence higher λ means less risk aversion.

(b) The Negative Exponential family of utility functions is parameterised by a constant absolute risk-aversion function $A(w) = A$, as follows:

$$U_{N(A)}(w) = -\exp(-Aw), \text{ where } A > 0. \tag{8}$$

Clearly, a higher value of A implies more risk aversion.

2.4 *Estimates of Absolute and Relative Risk-aversion*

To parameterise these utility functions, we need estimates of absolute or relative risk-version coefficients. Eisenhauer & Ventura (2003) pointed out that past research was inconclusive; estimates of average relative risk-aversion coefficients ranged from less than 1 to well over 40. Hoy & Witt (2007) illustrated their model using iso-elastic utilities with $R = 0.5$, 1 and 3. We will adopt a similar strategy, as follows.

Eisenhauer & Ventura (2003) estimated the risk-aversion function based on a thought experiment conducted by the Bank of Italy for its 1995 Survey of Italian Households'

Income and Wealth. Under certain assumptions, they estimated that a person with an average annual income of 46.7777 million lira had absolute risk-aversion coefficient 0.1837, and relative risk-aversion coefficient 8.59. (Guiso & Paiella (2006), based on the same study, estimated the relative risk aversion coefficient to be 1.92 for the $10th$ percentile and 13.25 for the $90th$ percentile.)

Allowing for the sterling/lira exchange rate in 1995 (average £1 = 2570.60 lira `http://fx.sauder.ubc.ca/`) and price inflation in the UK between July 1995 and June 2006 (Retail Price Index 149.1 and 198.5, respectively) an average income of 46.7777 million lira in 1995 equates to about £24,226 in 2006, not very different from the actual average of £25,810 (Jones (2005)).

We need utility functions of wealth, so an estimate of the wealth-income ratio is required. Estimates of this ratio in the literature are quite varied. According to H.M. Treasury (2005) in the U.K., it varies between 5 and 7 for total wealth, and between 2 and 4 for net financial wealth.

The Inland Revenue in the U.K. also publishes figures on personal wealth distribution `http://www.hmrc.gov.uk/stats/personal_wealth/menu.htm`. Their latest figure (for 2003) shows that 53% of the population has less than £50,000 and 83% has less than £100,000. As the distribution of wealth is positively skewed, we will assume a total wealth of $W = £100,000$. This gives a wealth-income ratio of 4 which is consistent with the figures published by H.M. Treasury (2005).

(a) The absolute risk-aversion function depends on the unit of wealth. Given utility functions $U(w)$ and $V(w)$ related by $U(cw) = V(w)$ for some constant c, their absolute risk-aversion functions are related by $A_U(cw) = A_V(w)/c$. Using exchange and inflation rates above, we suppose that a Briton in 2006 has absolute risk-aversion coefficient $8.967 \times 10^{-5} \approx 9 \times 10^{-5}$, denominated in 2006 pounds.

(b) The relative risk-aversion function does not depend on the unit of wealth and so the estimate of 8.59 can be used without any adjustment. We will use a rounded-off value of 9 in the remainder of the paper.

The formulation of utility functions with non-constant relative risk-aversion is an active area of research. Meyer & Meyer (2005) specified a form of marginal utility function which gives decreasing relative risk-aversion. Xie (2000) proposed a power risk-aversion utility function which can produce increasing, constant or decreasing risk-aversion depending on its parameterisation. These specialised utility functions are not yet in widespread use and we will not consider them further.

We will use the following utility functions for the purposes of illustration:

(a) Iso-elastic utilities with parameter $\lambda = 0.5$, 0 and -8, which corresponds to constant relative risk-aversion of 0.5, 1 and 9 respectively.

(b) Negative exponential utility with absolute risk-aversion coefficient $A = 9 \times 10^{-5}$.

Since iso-elastic utility with $\lambda = -8$ has absolute risk-aversion coefficient equal to 9×10^{-5} when wealth is £100,000, our assumption of $W = £100,000$ allows us to compare the two utility functions.
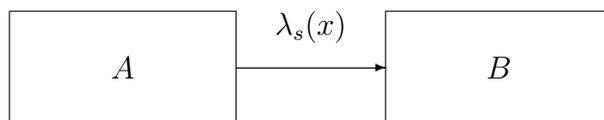
Figure 1: A two state model

3. A SIMPLE GENE-ENVIRONMENT INTERACTION MODEL

We will illustrate the principles of underwriting long-term insurance in the presence of a multifactorial disorder in the simple setting of the two-state continuous-time model in Figure 1. The insured event could be death or illness, and it is represented by an irreversible transition from state A to state B. The probability of transition is governed by the transition intensity $\lambda_s(x)$, which depends on age x, and the values of various risk factors which are labelled s (for 'stratum'). In essence, $\lambda_s(x)\,dx$ is the probability that a person who is healthy at age x should suffer the insured event during the next small time interval of length dx.

The risk factors arise from a 2×2 gene-environment interaction model. That is, there are two genotypes, denoted G and g, and two levels of environmental exposure, denoted E and e. We assume that G and E are adverse exposures while g and e are beneficial. Therefore, there are four risk groups or strata, that we label ge, gE, Ge and GE. Let the proportion of the population at a particular age (at which an insurance contract is sold) in stratum s be w_s. The epidemiology is defined as follows.
(a) We assume proportional hazards, so for each stratum s there is a constant k_s, independent of age, such that $\lambda_s(x)/\lambda_{ge}(x) = k_s$ for all ages x. Clearly $k_{ge} = 1$, and $k_s > 1$ for $s \neq ge$.
(b) We assume symmetry between genetic and environmental risks, as follows:
    (1) The probability of possessing the beneficial gene g is the same as the probability of exposure to the beneficial environment e, each denoted ω. Assuming independence, $w_{ge} = \omega^2$, $w_{gE} = w_{Ge} = \omega(1 - \omega)$ and $w_{GE} = (1 - \omega)^2$.
    (2) We assume that $k_{gE} = k_{Ge} = k$.
(c) The gene-environment interaction is represented by either an additive or a multiplicative model, as follows:
    (1) Additive Model: $k_{GE} = k_{Ge} + k_{gE} - k_{ge} = 2k - 1$.
    (2) Multiplicative Model: $k_{GE} = k_{Ge}k_{gE}/k_{ge} = k^2$.
    See Woodward (1999) for a discussion of additive and multiplicative models.

Therefore, the epidemiology is fully defined by the parameters $\lambda_{ge}(x)$, ω and k along with the choice of interaction.

This model could also be used to represent other forms of interacting risk factors, such as fixed, non-modifiable influences on the genetic risk. For example, instead of environment, e and E could represent maternal and paternal transmission, respectively, of the gene responsible for Huntington's disease. As economic modelling of multifactorial disorders advances from hypothetical to actual cases, the most distinctive feature of environmental factors may be that individuals can choose to modify them.

4. Insurance Premiums

4.1 *Single Premiums*

For simplicity, let the force of interest be $\delta = 0$. (This is consistent with the assumptions of Doherty & Thistle (1996), Hoy & Polborn (2000) and Hoy & Witt (2007).) Then the single premium for an insurance contract of term t years, with sum assured £1, sold to a person aged x who belongs to stratum s is:

$$q_s = 1 - \exp\left[-\int_0^t \lambda_s(x + y)dy\right] = 1 - (1 - q_{ge})^{k_s}. \tag{9}$$

If the proportion of insurance purchasers aged x is the same as the proportion in the population, w_s (for example if the stratum is not known to applicants or to insurers) observation of claim statistics will lead the insurer to charge a weighted average premium rate $\bar{q} = \sum_s w_s q_s = \sum_s w_s[1 - (1 - q_{ge})^{k_s}]$ per unit sum assured. Given our assumption that the k_s can all be expressed as simple functions of k, the stratum-specific and average premium rates can also be expressed as $q_s(k)$ and $\bar{q}(k)$.

4.2 *Threshold Premium*

Suppose all individuals have initial wealth W and that the net effect of suffering the insured event in the next n years is a loss of L. Define the loss ratio $f = L/W$. If no-one knows to which stratum they belong everyone will be willing to pay a single premium of up to:

$$P^* = W - U^{-1}[\bar{q}(k)U(W - L) + (1 - \bar{q}(k))U(W)]. \tag{10}$$

However, someone who knows they are in stratum s will be willing to pay a single premium of up to:

$$P_s^* = W - U^{-1}[q_s(k)U(W - L) + (1 - q_s(k))U(W)]. \tag{11}$$

P_s^* is smallest for stratum ge. So if the insurer, ignorant of the stratum, continues to charge premium $\bar{q}(k)L$, adverse selection will first appear if $\bar{q}(k)L > P_{ge}^*$. That is, if:

$$U(W - \bar{q}(k)L) < q_{ge}(k)U(W - L) + (1 - q_{ge}(k))U(W). \tag{12}$$

To be ignorant of the stratum in which an applicant exists, the insurer must be unable to observe *both* genotype and environment. In practice, the insurer may have partial knowledge, even if regulations bar the use of genetic information, because important environmental risk factors such as smoking may be freely observable.

4.3 *The Additive Epidemiological Model*

Replace the inequality in Equation 12 with an equality and solve for k; this represents the relative risk (of each risk factor) with respect to stratum ge, above which persons who know they are in stratum ge will cease to buy insurance. Doing this with iso-elastic utility with $\lambda \neq 0$ we obtain:

Table 1: The relative risk k above which persons in stratum ge with initial wealth $W = £100,000$ will not buy insurance, using $\omega = 0.5$ and an additive model.

Utility Function	q_{ge}	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	0.1	1.025	1.053	1.085	1.122	1.165	1.217	1.284	1.373	1.513
	0.2	1.024	1.050	1.081	1.116	1.158	1.209	1.274	1.364	1.506
$I(0.5)$	0.3	1.022	1.047	1.076	1.110	1.150	1.200	1.263	1.352	1.497
	0.4	1.021	1.044	1.072	1.103	1.142	1.189	1.251	1.339	1.486
	0.5	1.019	1.041	1.066	1.096	1.132	1.178	1.238	1.324	1.472
	0.1	1.051	1.110	1.180	1.264	1.368	1.504	1.691	1.976	2.524
	0.2	1.048	1.104	1.170	1.250	1.350	1.479	1.659	1.939	2.488
Log	0.3	1.045	1.098	1.160	1.235	1.330	1.453	1.626	1.898	2.451
	0.4	1.042	1.091	1.149	1.220	1.308	1.425	1.590	1.854	2.413
	0.5	1.039	1.084	1.138	1.203	1.286	1.395	1.551	1.805	2.372
	0.1	1.598	2.755	4.947	8.831	15.950	–	–	–	–
	0.2	1.546	2.512	4.153	6.972	14.430	–	–	–	–
$I(-8)$	0.3	1.498	2.322	3.664	6.148	–	–	–	–	–
	0.4	1.451	2.163	3.313	5.810	–	–	–	–	–
	0.5	1.405	2.023	3.035	6.107	–	–	–	–	–
	0.1	1.566	2.504	3.917	5.793	8.036	10.574	13.428	16.739	20.862
	0.2	1.516	2.292	3.337	4.617	6.119	7.911	10.226	13.900	–
$N(9e\text{-}5)$	0.3	1.468	2.126	2.963	3.972	5.204	6.857	9.812	–	–
	0.4	1.423	1.987	2.684	3.536	4.655	6.519	–	–	–
	0.5	1.379	1.864	2.457	3.206	4.305	7.636	–	–	–

$$(1 - \bar{q}(k)f)^\lambda = q_{ge}(1-f)^\lambda + (1 - q_{ge}). \tag{13}$$

In the special case of logarithmic utility (iso-elastic utility with $\lambda = 0$) we obtain:

$$1 - \bar{q}(k)f = (1-f)^{q_{ge}} \tag{14}$$

and under negative exponential utility:

$$e^{\bar{q}(k)AL} = q_{ge}e^{AL} + (1 - q_{ge}) \tag{15}$$

in which wealth W does not appear. Using $\omega = 0.5$ (a uniform distribution across strata) and an additive model, we solve Equations 13, 14 and 15 for k, given certain values of baseline risk q_{ge} and loss L, assuming an initial wealth of $W = £100,000$. The results are in Table 1. We observe the following:

(a) For low loss ratios, even small relative risks k will cause people in the baseline stratum to opt against insurance. This is as expected as small losses are relatively tolerable.

(b) As the loss ratio f increases, so does the relative risk at which adverse selection appears. This is simply risk aversion at work.

(c) The higher the baseline risk q_{ge} for a given loss ratio f, the lower the relative risk at which adverse selection appears.

(d) Lower risk-aversion, under iso-elastic utility, $(\lambda = 0.5)$ means that smaller relative risks would discourage members of the baseline stratum to buy insurance at the average premium, and for higher risk-aversion $(\lambda = -8)$ the reverse is true.

Comparing iso-elastic and negative exponential utilities, we see that the limiting relative risks are broadly similar for smaller losses. For larger losses, however, iso-elastic utility functions have much greater limiting relative risks. This is because risk-aversion increases as wealth falls under iso-elastic utility, while for negative exponential utility it is constant. As the fair actuarial premium for bigger losses increases and depletes wealth, risk-aversion under iso-elastic utility climbs above that under negative exponential utility, with the result shown.

4.4 *Immunity From Adverse Selection*

The missing entries in Table 1 mean that adverse selection never appears, whatever the relative risk k. Clearly, this must be related to the size of the high-risk strata, and their ability, or otherwise, to move the average premium enough to affect the baseline stratum. We may ask: given q_{ge} and f, is there some proportion w_{ge} in the lowest risk stratum above which members of that stratum will always buy insurance at the average premium rate? Begin by noting that:

$$\lim_{k\to\infty} \bar{q}(k) = \lim_{k\to\infty} \sum_s w_s[1 - (1 - q_{ge})^{k_s}] = w_{ge}q_{ge} + \sum_{s\neq ge} w_s = 1 - w_{ge}(1 - q_{ge}) \quad (16)$$

and that this limit is not a function of the k_s and thus holds for additive and multiplicative models. Substituting this limiting value in Equations 13 to 15, we can solve for w_{ge} as follows, for iso-elastic utility with $\lambda \neq 0$:

$$w_{ge} = \frac{1}{1 - q_{ge}}\left[1 - \frac{1 - (q_{ge}(1 - f)^{\lambda} + (1 - q_{ge}))^{1/\lambda}}{f}\right], \quad (17)$$

for logarithmic utility:

$$w_{ge} = \frac{1}{1 - q_{ge}}\left[1 - \frac{1 - (1 - f)^{q_{ge}}}{f}\right] \quad (18)$$

and for negative exponential utility:

$$w_{ge} = \frac{1}{1 - q_{ge}}\left[1 - \frac{\log[q_{ge}e^{AL} + (1 - q_{ge})]}{AL}\right]. \quad (19)$$

Values of $\omega = w_{ge}^{1/2}$ are given in Table 2. Values of $\omega < 0.5$ in Table 2 correspond to missing entries in Table 1. Table 2 shows just how uncommon an adverse exposure has to be to avoid adverse selection.

Table 2: The proportions ω exposed to each low-risk factor above which persons in the baseline stratum will buy insurance at the average premium regardless of the relative risk k, using different utility functions.

Utility Function	q_{ge}	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	0.1	0.999	0.997	0.996	0.994	0.991	0.989	0.985	0.981	0.974
	0.2	0.997	0.994	0.991	0.987	0.983	0.977	0.970	0.961	0.947
$I(0.5)$	0.3	0.996	0.992	0.987	0.981	0.974	0.966	0.955	0.941	0.919
	0.4	0.995	0.989	0.982	0.974	0.965	0.954	0.940	0.920	0.890
	0.5	0.993	0.986	0.978	0.968	0.956	0.942	0.924	0.899	0.860
	0.1	0.997	0.994	0.991	0.986	0.981	0.974	0.965	0.951	0.926
	0.2	0.995	0.989	0.981	0.973	0.962	0.949	0.932	0.906	0.859
Log	0.3	0.992	0.983	0.972	0.960	0.945	0.925	0.900	0.863	0.798
	0.4	0.989	0.977	0.963	0.947	0.927	0.902	0.870	0.823	0.743
	0.5	0.987	0.972	0.954	0.934	0.910	0.880	0.841	0.786	0.693
	0.1	0.969	0.916	0.830	0.719	0.603	0.496	0.398	0.304	0.203
	0.2	0.943	0.857	0.747	0.632	0.525	0.431	0.345	0.264	0.176
$I(-8)$	0.3	0.919	0.812	0.693	0.580	0.480	0.393	0.315	0.241	0.161
	0.4	0.897	0.776	0.653	0.543	0.448	0.367	0.294	0.225	0.150
	0.5	0.878	0.746	0.622	0.515	0.424	0.347	0.279	0.213	0.142
	0.1	0.971	0.927	0.868	0.802	0.738	0.682	0.635	0.595	0.562
	0.2	0.946	0.875	0.797	0.723	0.660	0.607	0.564	0.528	0.498
$N(9e\text{-}5)$	0.3	0.923	0.835	0.748	0.673	0.612	0.562	0.522	0.488	0.461
	0.4	0.903	0.802	0.712	0.637	0.577	0.530	0.492	0.460	0.434
	0.5	0.884	0.775	0.682	0.608	0.551	0.505	0.468	0.439	0.414

Table 3: The relative risk k above which persons in stratum ge with initial wealth $W = £100,000$ will not buy insurance, using $\omega = 0.9$ and an additive model.

Utility Function	q_{ge}	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	0.1	1.126	1.269	1.433	1.625	1.855	2.140	2.511	3.033	3.899
	0.2	1.120	1.258	1.419	1.613	1.852	2.158	2.577	3.212	4.419
$I(0.5)$	0.3	1.113	1.246	1.404	1.599	1.847	2.180	2.668	3.502	5.689
	0.4	1.106	1.233	1.387	1.582	1.841	2.210	2.807	4.108	–
	0.5	1.099	1.218	1.367	1.562	1.833	2.250	3.055	–	–
	0.1	1.257	1.563	1.934	2.399	3.004	3.839	5.101	7.368	13.841
	0.2	1.246	1.546	1.923	2.418	3.107	4.170	6.164	13.981	–
Log	0.3	1.233	1.526	1.910	2.444	3.268	4.844	–	–	–
	0.4	1.220	1.504	1.894	2.482	3.555	8.317	–	–	–
	0.5	1.205	1.479	1.876	2.542	4.296	–	–	–	–
	0.1	4.458	18.642	–	–	–	–	–	–	–
	0.2	4.823	–	–	–	–	–	–	–	–
$I(-8)$	0.3	5.705	–	–	–	–	–	–	–	–
	0.4	–	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–
	0.1	4.246	13.531	–	–	–	–	–	–	–
	0.2	4.514	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	0.3	5.109	–	–	–	–	–	–	–	–
	0.4	7.984	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–

Assuming $\omega = 0.5$ is perhaps extreme; it means that half the population possess a significant genetic risk factor (modulated by environment) yet to be discovered. This is by no means impossible, but we might expect most as-yet unknown risk factors to affect a smaller proportion of the population, simply because they are as-yet unknown. So, we increase ω to 0.9, so that only 10% of individuals are exposed to the adverse environment or possess the adverse gene. The relative risks k at which adverse selection appears are given in Table 3. They are larger than in Table 1 because the relative risk experienced by the smaller number of high-risk individuals has to be much higher to have the same impact on the average premium.

4.5 *The Multiplicative Epidemiological Model*

Table 4 shows relative risks above which adverse selection appears, assuming $\omega = 0.9$ and a multiplicative model. They can be compared with the values in Table 3. We observe the following:

(a) The missing entries are the same as in the additive model. This is because the limiting values of $\bar{q}(k)$ and ω do not depend on the model structure.

(b) The relative risk in stratum GE is higher in the multiplicative model ($k^2 > 2k - 1$) so persons in the baseline stratum will be less tolerant towards any given value of k.

Table 4: The relative risk k above which persons in stratum ge with initial wealth $W = £100,000$ will not buy insurance, using $\omega = 0.9$ and a multiplicative model.

Utility Function	q_{ge}	10	20	30	40	50	60	70	80	90
						loss L in £'000				
	0.1	1.125	1.265	1.424	1.608	1.825	2.090	2.431	2.907	3.701
	0.2	1.119	1.255	1.412	1.598	1.825	2.115	2.511	3.119	4.315
$I(0.5)$	0.3	1.113	1.243	1.398	1.586	1.824	2.144	2.617	3.447	5.660
	0.4	1.106	1.230	1.381	1.571	1.822	2.181	2.773	4.086	–
	0.5	1.098	1.216	1.362	1.553	1.817	2.229	3.037	–	–
	0.1	1.254	1.549	1.899	2.328	2.880	3.645	4.839	7.107	13.706
	0.2	1.243	1.533	1.892	2.360	3.018	4.065	6.086	13.967	–
Log	0.3	1.231	1.516	1.884	2.399	3.212	4.805	–	–	–
	0.4	1.218	1.495	1.873	2.449	3.527	8.314	–	–	–
	0.5	1.203	1.472	1.859	2.521	4.288	–	–	–	–
	0.1	4.223	18.561	–	–	–	–	–	–	–
	0.2	4.723	–	–	–	–	–	–	–	–
$I(-8)$	0.3	5.676	–	–	–	–	–	–	–	–
	0.4	–	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–
	0.1	4.024	13.391	–	–	–	–	–	–	–
	0.2	4.410	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	0.3	5.073	–	–	–	–	–	–	–	–
	0.4	7.981	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–

This is why the values in Table 4 are smaller than those in Table 3.

(c) However the differences between the additive and multiplicative models are not very large. If $k \approx 1$, then $k^2 \approx 2k - 1$, and for large values of ω (which arguably is most realistic) the impact of stratum GE is relatively small. In view of this, we will use only the additive model from now on.

4.6 *Loss* versus *Coverage*

Our simple model assumes that everyone risks the same loss L, and chooses to insure it 100% if they insure at all. A more realistic model, as pointed out by a referee, might assume that persons knowing themselves to be in stratum s choose insurance cover of C_s, not necessarily equal to L (indeed not necessarily bounded by L unless the insurer limits the coverage by reference to some objective measure of L or W). Assuming the insurer charges an average premium rate $\bar{q}(k)$ as before, C_s will be chosen to maximise $q_s U(W - L + C_s - \bar{q}(k)C_s) + (1 - q_s)U(W - \bar{q}(k)C_s)$. This extension of the model would be of interest in its own right; however some experiments (not shown) confirm that it does not change the qualitative nature of our conclusions.

4.7 *Genetic Information and Behaviour*

The introduction of new genetic information — ability to learn one's own genotype — may lead high-risk people in particular to alter their behaviour to ameliorate the risk. Thus the composition of the risk groups may not be the same before and after genetic testing (say) becomes available. This possibility is more plausible for multifactorial diseases than for single-gene disorders, since there will often be modifiable environmental or lifestyle interactions. For example, if our environmental variable was E = 'smoker' and e = 'non-smoker', persons initially in stratum GE might be particularly likely to stop smoking, and (perhaps after some time) move to stratum Ge. The low-risk strata will be enlarged, which will: (a) cause the weighted average premium to fall; and (b) as in Table 2, make it more likely that low-risk individuals will buy insurance regardless of the relative risks. Therefore, our results err on the pessimistic side.

Such behavioural effects can, in principle, be modelled by allowing transitions between strata, after genetic testing and before insurance is purchased. For example, suppose $\omega = 0.9$ and 1% of the population is initially in stratum GE, but that after genetic tests become available, half of those in stratum GE move to stratum Ge. Table 3 shows the relative risk thresholds before, and Table 5 after, the introduction of genetic tests. There is an appreciable difference, even though only 0.5% of the population has changed its behaviour. However, since we have no greater insight than this into *how* behaviour might change, we interpret all our results except those in Table 5 as being after any behavioural changes have taken effect. When real epidemiological studies eventually become available, the effect of modified behaviour should not be overlooked.

5. Critical Illness Insurance

5.1 *A Heart Attack Model*

We now model the specific example of CI insurance. We will focus on heart attack risk, building upon two earlier papers, in which the reader can find full details.

Table 5: The relative risk k above which persons in stratum ge with initial wealth $W = £100,000$ will not buy insurance, using $w_{ge} = 0.81$, $w_{Ge} = 0.095$, $w_{gE} = 0.09$ and $w_{GE} = 0.005$ and an additive model.

Utility Function	q_{ge}	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	0.1	1.167	1.356	1.571	1.821	2.118	2.482	2.949	3.597	4.649
	0.2	1.159	1.339	1.548	1.794	2.092	2.468	2.971	3.714	5.078
$I(0.5)$	0.3	1.150	1.321	1.523	1.765	2.067	2.462	3.023	3.951	6.272
	0.4	1.140	1.302	1.496	1.734	2.040	2.463	3.125	4.508	–
	0.5	1.129	1.282	1.466	1.699	2.012	2.478	3.339	–	–
	0.1	1.341	1.741	2.219	2.808	3.561	4.576	6.071	8.662	15.674
	0.2	1.324	1.709	2.180	2.781	3.593	4.801	6.974	15.032	–
Log	0.3	1.306	1.675	2.142	2.767	3.693	5.389	–	–	–
	0.4	1.286	1.640	2.102	2.768	3.928	8.788	–	–	–
	0.5	1.265	1.601	2.061	2.794	4.621	–	–	–	–
	0.1	5.315	20.677	–	–	–	–	–	–	–
	0.2	5.523	–	–	–	–	–	–	–	–
$I(-8)$	0.3	6.288	–	–	–	–	–	–	–	–
	0.4	–	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–
	0.1	5.063	15.347	–	–	–	–	–	–	–
	0.2	5.182	–	–	–	–	–	–	–	–
$N(9e-5)$	0.3	5.666	–	–	–	–	–	–	–	–
	0.4	8.454	–	–	–	–	–	–	–	–
	0.5	–	–	–	–	–	–	–	–	–

(a) Gutiérrez & Macdonald (2003) parameterised the CI model shown in Figure 2, using medical studies and population data. Therefore, in particular, $\lambda_{12}(x)$ denotes the rate of onset of heart attacks in the general population (different for males and females).

(b) Macdonald, Pritchard & Tapadar (2006) assumed that a 2×2 gene-environment interaction affected heart attack risk, with genotypes G and g, and environmental exposures E and e, upper case representing higher risk. So there are four strata for each sex — ge, gE, Ge and GE. The authors showed that it is possible to hypothecate assumptions on strata-specific relative risks, in a way which is consistent with the rate of onset in the general population. We will use a similar technique here.
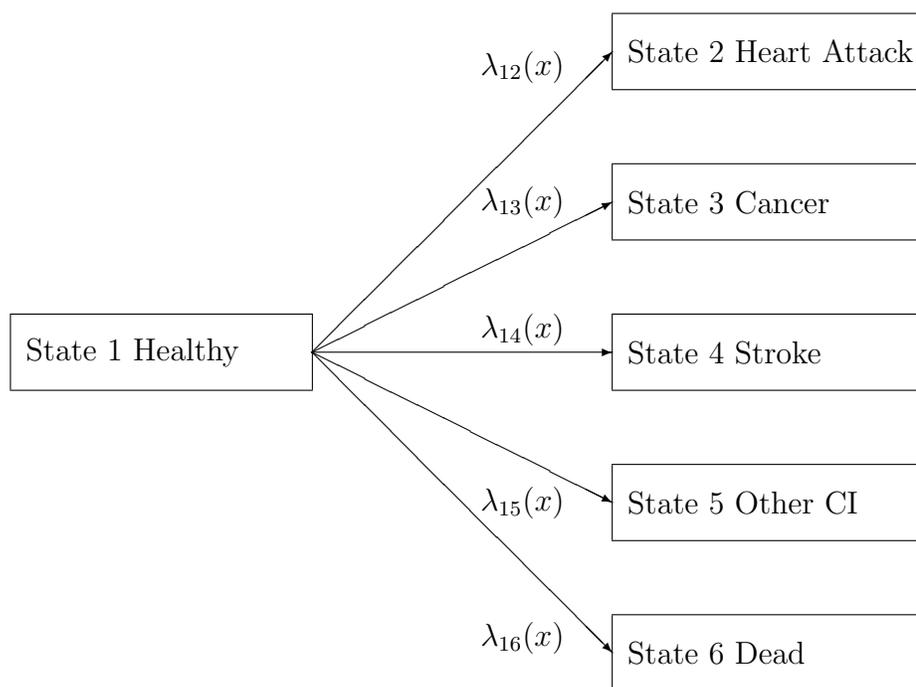

Figure 2: A full critical illness model.

Consider all healthy individuals aged x. If $\bar{q}$ denotes the probability that a healthy person aged x has a heart attack before age $x + t$, it can be calculated from the heart attack transition intensity of the general population as follows:

$$\bar{q} = 1 - \exp\left[-\int_0^t \lambda_{12}(x + y)dy\right] \tag{20}$$

Now, for males and females separately, let c denote the relative risk in the baseline stratum ge with respect to the general population, and let k_s denote the relative risk in stratum s with respect to stratum ge, in both cases assumed to be constant at all ages
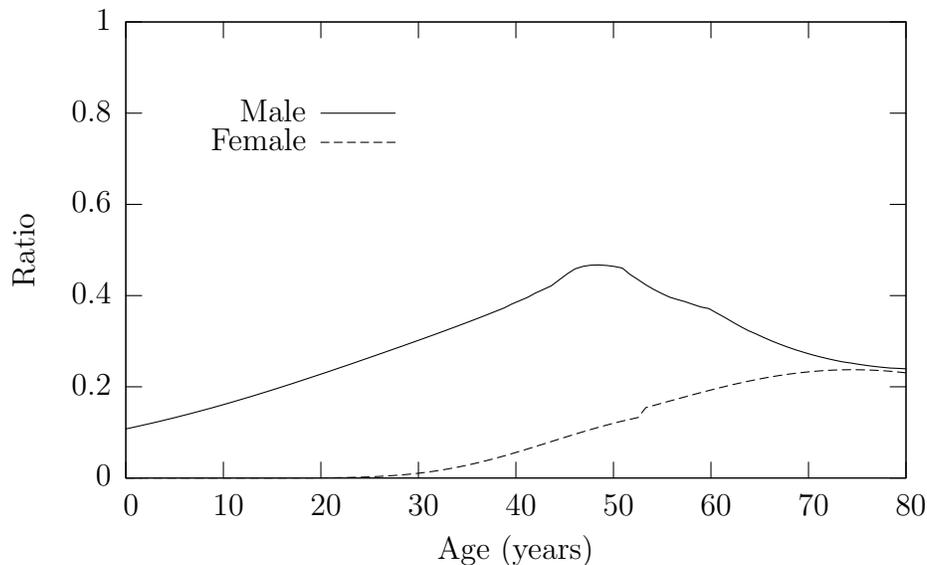
Figure 3: The ratio of heart attack transition intensity to total critical illness transition intensity, by gender.

(in other words, we assume a proportional hazards model). If we denote the rate of onset of heart attack in stratum s by $\lambda_{12}^s(x)$, it is given by:

$$\lambda_{12}^s(x) = c \times k_s \times \lambda_{12}(x). \tag{21}$$

Suppose that at age, x, the proportion of healthy individuals who are in stratum s is w_s. In stratum s, let q_s be the probability that a healthy person age x has a first heart attack before reaching age $x + t$. Then using Equations 20 and 21, we can show that:

$$q_s = 1 - \exp\left[-\int_0^t \lambda_{12}^s(x+y)dy\right] = 1 - (1 - \bar{q})^{ck_s}. \tag{22}$$

Equating the weighted average probability over all strata with the population probability, that is, $\bar{q} = \sum w_s q_s$, we have:

$$\bar{q} = \sum w_s[1 - (1 - \bar{q})^{ck_s}]. \tag{23}$$

Given the relative risks, the population proportions and the estimated $\lambda_{12}(x)$, we can solve this for c, which fully specifies the stratum-specific intensities $\lambda_{12}^s(x)$.

5.2 *Threshold Premium for Critical Illness Insurance*

To extend the two-state insurance model of Section 3 to the CI model with six states, we make some simplifying assumptions.

(a) We will model gene-environment interactions affecting heart attack risk alone, leaving other intensities unaffected. This is not completely realistic, since many known risk factors for heart disease are also risk factors for other disorders.

Table 6: The premium rates of critical illness contracts of duration 15 years.

Age	Male	Female
25	0.013787	0.018746
35	0.048413	0.049715
45	0.136363	0.110434

(b) The heart attack transition intensity is different for males and females. Figure 3 shows the ratio $\lambda_{12}(x)/\sum_{j=2}^{5}\lambda_{1j}(x)$ for both sexes. Heart attack is the predominant CI among middle-aged men, while among women, heart attack is increasingly prominent from age 30 onwards, but cancer is the dominant CI at all ages. The ratio for males stays significantly higher than the ratio for females, except at very high ages. Hence we might expect adverse selection to appear at different relative risk thresholds for the two sexes.

5.3 *Premium Rates for Critical Illness Insurance*

As examples, we model single-premium CI insurance contracts of duration 15 years sold to males and females aged 25, 35 and 45. First, assuming all transition intensities are as given in Gutiérrez & Macdonald (2003), we compute the single premiums as expected present values (EPVs) of the benefit payments by solving Thiele's differential equations (see Norberg (1995)) numerically. Again for simplicity, we assume the force of interest $\delta = 0$. Table 6 gives the CI premium rates per unit sum assured for these contracts.

We make the same epidemiological assumptions as before, namely that $k_{gE} = k_{Ge} = k$; that an additive model ($k_{GE} = 2k - 1$) applies, and that $w_{ge} = \omega^2$, $w_{gE} = w_{Ge} = \omega(1 - \omega)$, and $w_{GE} = (1 - \omega)^2$, where $\omega = 0.9$ (the more realistic assumption); and also that initial wealth is $W = £100,000$. Given the relative risks, we obtain c and hence the the heart attack intensity for each sex and stratum as in Section 5.1. This allows us to calculate stratum-specific premium rates.

Let P_s denote the single premium rate for unit CI insurance in stratum s. Note that apart from the stratum-specific heart attack risk, P_s also covers the risk of all other CIs, which are assumed to be the same for all strata. Let $\bar{P}$ denote the population average premium rate for unit CI insurance (the averaging being over all strata for a given gender). As before, since we are ignoring interest rates and profit margins, the various premium rates defined above are the same as the probabilities of the event insured against. Then define a function $Z(P)$ of a premium P as follows:

$$Z(P) = U(W - \bar{P}L) - [PU(W - L) + (1 - P)U(W)]. \tag{24}$$

Table 7: $P^{\dagger}$ for males, which solves $Z(P) = 0$, for different combinations of utility functions and losses, using initial wealth W = £100,000.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	0.013438	0.013068	0.012674	0.012250	0.011788	0.011277	0.010695	0.010004	0.009102
$I(0.5)$	35	0.047229	0.045969	0.044622	0.043167	0.041577	0.039808	0.037788	0.035378	0.032216
	45	0.133321	0.130058	0.126534	0.122691	0.118448	0.113678	0.108172	0.101522	0.092679
	25	0.013095	0.012374	0.011620	0.010826	0.009980	0.009065	0.008055	0.006891	0.005423
Log	35	0.046062	0.043604	0.041019	0.038282	0.035353	0.032171	0.028636	0.024543	0.019348
	45	0.130316	0.123918	0.117108	0.109801	0.101879	0.093158	0.083326	0.071772	0.056865
	25	0.008388	0.004503	0.002062	0.000773	0.000223	0.000045	0.000005	0.000000	0.000000
$I(-8)$	35	0.029922	0.016319	0.007596	0.002893	0.000849	0.000174	0.000021	0.000001	0.000000
	45	0.087752	0.049912	0.024272	0.009674	0.002978	0.000642	0.000081	0.000004	0.000000
	25	0.008554	0.004976	0.002733	0.001429	0.000719	0.000351	0.000167	0.000078	0.000036
$N(9e\text{-}5)$	35	0.030512	0.018032	0.010061	0.005349	0.002734	0.001356	0.000656	0.000312	0.000146
	45	0.089459	0.055093	0.032069	0.017804	0.009517	0.004938	0.002504	0.001247	0.000613

Table 8: $P^\dagger$ for females, which solves $Z(P) = 0$, for different combinations of utility functions and losses, using initial wealth $W = £100,000$.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	0.018273	0.017773	0.017239	0.016664	0.016038	0.015344	0.014554	0.013616	0.012389
$I(0.5)$	35	0.048500	0.047209	0.045827	0.044334	0.042702	0.040886	0.038813	0.036339	0.033093
	45	0.107899	0.105188	0.102269	0.099094	0.095600	0.091684	0.087179	0.081758	0.074580
	25	0.017809	0.016833	0.015811	0.014734	0.013586	0.012344	0.010971	0.009388	0.007390
Log	35	0.047304	0.044782	0.042131	0.039322	0.036315	0.033050	0.029420	0.025217	0.019880
	45	0.105398	0.100089	0.094460	0.088443	0.081945	0.074821	0.066825	0.057471	0.045463
	25	0.011431	0.006150	0.002823	0.001060	0.000307	0.000062	0.000007	0.000000	0.000000
$I(-8)$	35	0.030745	0.016778	0.007814	0.002978	0.000875	0.000180	0.000021	0.000001	0.000000
	45	0.070219	0.039438	0.018924	0.007438	0.002256	0.000479	0.000059	0.000003	0.000000
	25	0.011657	0.006796	0.003740	0.001961	0.000989	0.000483	0.000231	0.000108	0.000050
$N(9e\text{-}5)$	35	0.031351	0.018539	0.010350	0.005506	0.002817	0.001397	0.000677	0.000322	0.000151
	45	0.071593	0.043550	0.025029	0.013714	0.007231	0.003700	0.001849	0.000908	0.000439

Table 9: The population average premium rate for CI insurance, P_0, as if heart attack risk were absent ($\lambda_{12} = 0$).

Age	Male	Female
25	0.009821	0.018326
35	0.031290	0.046485
45	0.092818	0.097947

Note that $Z(P_{ge}) < 0$ is the condition under which adverse selection will appear, equivalent to Equation 12 of Section 4.2. Or, let $P^\dagger$ be the solution of $Z(P) = 0$. Then $P_{ge} < P^\dagger$ is the condition for adverse selection to appear. Tables 7 and 8 show $P^\dagger$ for males and females respectively. It depends on the utility function but not on the epidemiological model. For the 2-state model, Equation 12 was central in our analysis. Given: (a) a model structure (additive or multiplicative), the baseline risk q_{ge}, and the proportion ω with low values of each risk factor; and (b) noting that the average risk $\bar{q}$ was an increasing function of the relative risk parameter k; we obtained a minimum value of k for which adverse selection first appears.

We would like to do the same for the CI insurance model. However, there are important differences between the two models.

(a) In the 2-state model we specified the baseline risk and relative risks, and these determined the average risk. In the CI insurance model, we specify the average risk (given by the population heart attack risk) and the relative risks, and these determine the baseline risk, in the form of the relative risk c. Clearly increasing the relative risk k will cause c to fall, hence also the premium P_{ge}. To make this dependence clear, we will write $c(k)$ and $P_{ge}(k)$ in this section. It will also be useful to note that the probability q_{ge} of a heart attack similarly depends on k, and write $q_{ge}(k)$.

(b) However, unlike in the 2-state model, $P_{ge}(k)$ has a lower bound, denoted P_0, given by the population average premium rate for CI insurance as if heart attack risk were absent ($\lambda_{12} = 0$ and $c = 0$). These values are shown in Table 9. They do not depend on the epidemiological model or the utility function. Clearly $P_{ge}(k) \geq P_0$, no matter how high k becomes. Thus we have two possibilities: $\lim_{k\to\infty} P_{ge}(k) = P_0$ (equivalently $\lim_{k\to\infty} c(k) = 0$); or $\lim_{k\to\infty} P_{ge}(k) > P_0$ (equivalently $\lim_{k\to\infty} c(k) > 0$). We return to this point in Section 5.4.

(c) If $P_{ge}(k)$ is a strictly decreasing function, which it is for the utility functions we are using, adverse selection is possible if $\lim_{k\to\infty} P_{ge}(k) < P^\dagger$, and in such cases we can solve $P_{ge}(k) = P^\dagger$ for the threshold value of k above which adverse selection will appear. Tables 10 and 11 show these values for the various utility functions and loss levels, for males and females respectively. The missing values correspond to combinations of parameters such that $\lim_{k\to\infty} P_{ge}(k) > P^\dagger$, for which adverse selection will not appear.

(d) Another consequence of this is that there is a level of insured loss, that we denote L_0, above which adverse selection cannot occur, because fixing $L > L_0$ in Equation 24 and solving for $P^\dagger$ yields a solution $P^\dagger < P_{ge}(k)$ for all k. Table 12 gives the values of L_0, for the usual utility functions and initial wealth £100,000. The missing values

Table 10: The relative risk k above which males of different ages in stratum ge with initial wealth $W = £100,000$ will not buy critical illness insurance policies of term 15 years, where $\omega = 0.9$.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	1.484	2.111	2.960	4.183	6.117	9.698	18.869	105.569	–
$I(0.5)$	35	1.376	1.846	2.450	3.262	4.420	6.226	9.509	17.715	93.578
	45	1.389	1.886	2.544	3.456	4.808	7.027	11.388	24.239	–
	25	2.062	3.783	7.068	15.883	122.410	–	–	–	–
Log	35	1.808	2.998	4.917	8.530	17.855	98.596	–	–	–
	45	1.843	3.138	5.339	9.794	23.063	765.192	–	–	–
	25	–	–	–	–	–	–	–	–	–
$I(-8)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–

Table 11: The relative risk k above which females of different ages in stratum ge with initial wealth $W = £100,000$ will not buy critical illness insurance policies of term 15 years, where $\omega = 0.9$.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	–	–	–	–	–	–	–	–	–
$I(0.5)$	35	4.031	18.470	–	–	–	–	–	–	–
	45	2.293	4.710	10.770	52.668	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
Log	35	15.856	–	–	–	–	–	–	–	–
	45	4.459	26.155	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$I(-8.0)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–

Table 12: The loss L_0 in £,000 above which adverse selection cannot occur. Initial wealth $W = $ £100,000.

| Gender | Age | Utility Function | | | |
		$I(0.5)$	Log	$I(-8)$	$N($9e-5$)$
	25	82.3	51.8	7.1	7.2
Male	35	92.3	62.6	9.2	9.5
	45	89.9	60.4	8.9	9.2
	25	8.9	4.5	0.5	0.5
Female	35	25.3	13.3	1.5	1.6
	45	43.4	23.9	2.9	2.9

in Tables 10 and 11 occur for losses $L > L_0$.

The general pattern of threshold relative risks for males given in Table 10 is similar to that in Section 4; what is of most interest are their absolute values, since we have tried to suggest plausible models for both the risk model and the utility functions.
(a) For iso-elastic utility with $\lambda = -8$ and negative exponential utility with parameter $A = 9 \times 10^{-5}$, we find no evidence at all of adverse selection.
(b) For all utility functions and at all loss levels, if adverse selection can appear, it does so at higher levels of relative risk than under the two-state model. This is because the impact of the gene and environment on heart attack risk is diluted by the presence of the other CIs, whose risks are assumed to be independent of the g/G genotypes and e/E environments. Only for the lowest levels of loss are these relative risks in the range that might be typical of relatively common multifactorial disorders; by definition, we do not expect studies like UK Biobank to lead to the discovery of hitherto unknown high risk genotypes.
(c) When adverse selection can appear, the relative risk threshold first decreases and then increases with age. This is because among CIs the importance of heart attack peaks at around age 45 as can be seen from Figure 6.

The threshold relative risks for females are given in Table 11. We observe the following:
(a) The threshold relative risks are much higher than those for males, in all cases. This is because heart attacks form a smaller proportion of all CIs for females, so a larger increase in heart attack risk is needed to trigger adverse selection.
(b) As for males, at levels of absolute and relative risk-aversion that we regard as most plausible (consistent with the Bank of Italy study) we find no evidence that adverse selection is likely.
(c) In contrast to males, the threshold relative risks decrease with age. The reason is clear from Figure 3; for females the relative importance of heart attack increases with age.
(d) Adverse selection appears to be possible only for: (i) smaller losses; and (ii) extremely low levels of risk aversion.

5.4 *High Relative Risks*

In Section 4.4, we considered relative risks that increased without limit, for the simple 2-state insurance model. We saw that, even in this extreme case, if stratum ge was large enough, adverse selection would not appear. In this section, we consider high relative risks (of heart attack) in the CI insurance model.

We assume the heart attack rates in the general population $\lambda_{12}(x)$ are fixed at their estimated values (Gutiérrez & Macdonald, 2003). From Equation 23 we obtain:

$$
\begin{aligned}
1 - \bar{q} &= 1 - \sum_s w_s[1 - (1 - \bar{q})^{c(k)k_s}] \\
&= w_{ge}(1 - \bar{q})^c + \sum_{s \neq ge} w_s(1 - \bar{q})^{ck_s}.
\end{aligned}
\tag{25}
$$

Differentiation shows the right-hand side to be a decreasing function of c and of each k_s ($s \neq ge$), all other quantities held constant in each case. Also, if $c = 1$ the right-hand side is less than $(1 - \bar{q})$ while if $c = 0$ it is greater than $(1 - \bar{q})$. Hence, as we increase the k_s without limit, c must decrease, and being bounded below it must have a limit. The limit could be zero or non-zero. We can easily see that if c has a non-zero limit (necessarily positive) then the last term on the right-hand side of Equation 25 vanishes and the limit must be:

$$
\lim_{\substack{k_s \to \infty \\ s \neq ge}} c = 1 - \frac{\log w_{ge}}{\log(1 - \bar{q})}
\tag{26}
$$

which in turn implies $(1 - \bar{q}) < w_{ge}$. On the other hand if $(1 - \bar{q}) > w_{ge}$, then c cannot have non-zero limit, so the equation:

$$
\lim_{\substack{k_s \to \infty \\ s \neq ge}} \sum_{s \neq ge} w_s(1 - \bar{q})^{ck_s} = (1 - \bar{q}) - w_{ge}
\tag{27}
$$

holds. Since the left-hand side is finite, at least one of the products ck_s tends to a finite limit as the $k_s \to \infty$. However, we have not specified here how the quantities k_s ($s \neq ge$) jointly approach infinity, so the behaviour of c is not easy to analyse in general. It is greatly simplified if the k_s are simple functions of a single parameter k, which is the case in our assumed epidemiological model (in which case we again make explicit the dependence of c by writing $c(k)$). For example, under an additive model with symmetry between genetic and environmental risks, Equation 25 can be written as:

$$
\begin{aligned}
1 - \bar{q} &= \omega^2(1 - \bar{q})^c + 2\omega(1 - \omega)(1 - \bar{q})^{c(k)k} + (1 - \omega)^2(1 - \bar{q})^{c(k)(2k-1)} \\
&= (1 - \bar{q})^{c(k)}[\omega + (1 - \omega)(1 - \bar{q})^{c(k)(k-1)}]^2
\end{aligned}
\tag{28}
$$

therefore:

$$
k = 1 + \frac{\log[(1 - \bar{q})^{(1-c(k))/2} - \omega] - \log(1 - \omega)}{c(k) \log(1 - \bar{q})}.
\tag{29}
$$

Table 13: $\bar{q}$, the probability that a healthy person aged x has a heart attack before age $x + t$, for policy duration $t = 15$ years.

Age	Male	Female
25	0.004743	0.000541
35	0.021454	0.004299
45	0.059959	0.017616

If $\omega^2 > (1 - \bar{q})$ then as $k \to \infty$, the limiting value of $c(k)$ is non-zero. Otherwise, when $\omega^2 < (1 - \bar{q})$, $c(k) \to 0$, and Equation 29 yields the finite limiting value:

$$\lim_{k \to \infty} c(k)k = \frac{\log[(1 - \bar{q})^{1/2} - \omega] - \log(1 - \omega)}{\log(1 - \bar{q})}. \tag{30}$$

So, in summary:

$$\lim_{k \to \infty} c(k) = \begin{cases} 0 & \text{if } w_{ge} \leq (1 - \bar{q}) \\ 1 - \frac{\log w_{ge}}{\log(1 - \bar{q})} & \text{if } w_{ge} > (1 - \bar{q}). \end{cases} \tag{31}$$

We want to find out if the baseline stratum ge can ever be large enough that adverse selection will never appear, no matter how large k becomes. Hence we want to understand the behaviour of $\lim_{k \to \infty} P_{ge}(k)$ as a function of w_{ge}. Equation 31 shows that we must treat separately the cases $w_{ge} \leq (1 - \bar{q})$ and $w_{ge} > (1 - \bar{q})$. Values of $\bar{q}$ are given in Table 13. (Note that $P_0 + \bar{q} \neq \bar{P}$, because in a competing risks model removing one cause of decrement increases the probabilities of the other decrements occurring.)

(a) If $P_0 > P^\dagger$ the result is trivial, since $\lim_{k \to \infty} P_{ge}(k) \geq P_0$ for any value of w_{ge}, and adverse selection can never occur.

(b) If $P_0 < P^\dagger$ adverse selection will occur if $w_{ge} \leq (1 - \bar{q})$, since then $\lim_{k \to \infty} P_{ge}(k) = P_0$.

(c) The non-trivial case is $P_0 < P^\dagger$ and $w_{ge} > (1 - \bar{q})$, since then $\lim_{k \to \infty} P_{ge}(k) > P_0$. We can show that $\lim_{k \to \infty} P_{ge}(k)$ is an increasing function of w_{ge} in this range, because the limit of the heart attack probability $\lim_{k \to \infty} q_{ge}(k)$ is (use Equation 26 to write:

$$\lim_{k \to \infty} q_{ge}(k) = \lim_{k \to \infty} [1 - (1 - \bar{q})^{c(k)}] = 1 - \frac{(1 - \bar{q})}{w_{ge}} \tag{32}$$

and differentiate). The function $\lim_{k \to \infty} P_{ge}(k)$ is continuous and increases from P_0 to $\bar{P}$ as w_{ge} increases from $(1 - \bar{q})$ to 1, the upper limit being attained when all the strata have collapsed into one, and $c = 1$. Since $P^\dagger < \bar{P}$ for any concave utility function, the intermediate value theorem guarantees that there exists a unique value of w_{ge} such that $\lim_{k \to \infty} P_{ge}(k) = P^\dagger$; that is, such that adverse selection can never appear if w_{ge} exceeds this value.

Tables 14 and 15 give the threshold values of $\omega = w_{ge}^{1/2}$ above which no adverse selection takes place, in the additive model with gene-environment symmetry, for males and females respectively. Missing values indicate that adverse selection will never appear.

Table 14: The proportions ω exposed to each low-risk factor above which persons in the baseline stratum will buy insurance at the average premium regardless of the relative risk k, using different utility functions, for males purchasing CI insurance.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	1.000	1.000	0.999	0.999	0.999	0.998	0.998	0.998	–
$I(0.5)$	35	0.999	0.998	0.998	0.997	0.996	0.995	0.993	0.992	0.990
	45	0.998	0.995	0.993	0.990	0.987	0.984	0.980	0.975	–
	25	1.000	0.999	0.999	0.998	0.998	–	–	–	–
Log	35	0.999	0.997	0.995	0.994	0.992	0.990	–	–	–
	45	0.996	0.991	0.986	0.981	0.976	0.970	–	–	–
	25	–	–	–	–	–	–	–	–	–
$I(-8)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–

When it is possible, the threshold value of ω ranges from 0.970 to 1 for males and 0.992 to 0.999 for females. As the relative risks in Tables 10 and 11 are based on $\omega = 0.9$, this explains the missing values in those tables.

This pattern is quite unexpected. If adverse selection can occur, then a large enough baseline stratum does confer immunity from it, but it has to be very large indeed, all but a few percent of the population. But once the threshold is crossed, adverse selection cannot appear at all, even if very few people are in the baseline stratum. This had no counterpart in the 2-state model, and it is caused by the presence of substantial other risks not affected by the gene-environment variants.

6. Conclusions

Until now, genetical research on information asymmetry and adverse selection has taken one of two routes — models of single-gene disorders and work on the economic welfare effects of genetic testing. In this paper, we have represented multifactorial disorders using standard epidemiological models and analysed circumstances leading to adverse selection, taking economic factors into account in a simple way through expected utility.

We used a range of iso-elastic utilities (including the special case of logarithmic utility) and a negative exponential utility, to represent constant relative and absolute risk aversion, respectively. They were parameterised to be reasonably consistent with some estimates based on survey data, but also to allow comparability, given our chosen level of wealth of £100,000.

We used a simple 2×2 gene-environment interaction model, assuming that information on status within the model was available only to the consumers and not to the insurer. Competition leads insurers to charge actuarially fair premiums, based on expected losses

Table 15: The proportions ω exposed to each low-risk factor above which persons in the baseline stratum will buy insurance at the average premium regardless of the relative risk k, using different utility functions, for females purchasing CI insurance.

Utility Function	Age	loss L in £'000								
		10	20	30	40	50	60	70	80	90
	25	–	–	–	–	–	–	–	–	–
$I(0.5)$	35	0.999	0.998	–	–	–	–	–	–	–
	45	0.998	0.996	0.994	0.992	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
Log	35	0.998	–	–	–	–	–	–	–	–
	45	0.996	0.993	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$I(-8)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–
	25	–	–	–	–	–	–	–	–	–
$N(9e\text{-}5)$	35	–	–	–	–	–	–	–	–	–
	45	–	–	–	–	–	–	–	–	–

given the information they have. Adverse selection will not occur as long as members of the least risky stratum (who know their status) can still increase their expected utility by insuring at the average price.

First, we studied a simple 2-state insurance model, with constant relative risks in different risk strata defined by the gene-environment model. We found that adverse selection does not appear unless purchasers are not very risk averse, and insure only a small proportion of their wealth; or unless the elevated risks implied by genetic information are implausibly high, bearing in mind the nature of multifactorial risk. In many cases adverse selection is impossible if the low-risk stratum is large enough, these levels being quite compatible with plausible multifactorial disorders.

We applied the same gene-environment interaction model, assumed to affect the risk of heart attacks, to CI insurance. As heart-attack risk is just part of the risk of all CIs, the impact of the gene-environment risk factor was diluted, compared with the 2-state insurance model where the total risk was influenced. Our results showed complete absence of adverse selection at realistic risk-aversion levels, irrespective of the stratum-specific risks, for males and females. Moreover, the existence of risks other than of heart attack, and the constraint of differential heart-attack risk to be consistent with the average population risk, introduced a threshold effect absent from the 2-state model. When adverse selection was possible at all (low risk aversion, low loss ratios) only an unfeasibly high proportion of the population in the low-risk stratum would avoid it, but when the threshold was crossed adverse selection vanished no matter what the size of the low-risk stratum.

The results from both 2-state and CI insurance models suggest that in circumstances that are plausibly realistic, private genetic information, relating to multifactorial risks,

that is available only to customers does not lead to adverse selection. This conclusion is strongest in the more realistic CI insurance model.

We have not considered what might happen if insurers were allowed access to this genetic information. The opportunity would then exist to underwrite using that information. If one believed that social policy is best served by solidarity, the important question is whether insurers would find it worthwhile to use the genetic information. Further research would be useful, to investigate the costs of acquiring and interpreting genetic information relating to common diseases, compared with the benefits in terms of possibly more accurate risk classification, in both cases in the context of multifactorial risk.

Acknowledgements

References

ABEL, A. (1986). Capital accumulation and uncertain lifetimes with adverse selection. *Econometrica*, **54**, 1079–97.

BINMORE, K. (1991). *Fun and games: A text on game theory.* Houghton Mifflin.

BRUGIAVINI, A. (1993). Uncertainty resolution and the timing of annuity purchases. *Journal of Public Economics*, **50**, 31–62.

DOHERTY, N.A. & THISTLE, P.D. (1996). Adverse selection with endogeneous information in insurance markets. *Journal of Public Economics*, **63**, 83–102.

EISENHAUER, J.G. & VENTURA, L. (2003). Survey measures of risk aversion and prudence. *Applied Economics*, **35**, 1477–1484.

GUISO, L. & PAIELLA, M. (2006), The rôle of risk aversion in predicting individual behavior, in *Competitive Failures in Insurance Markets: Theory and Policy Implications*, eds. Chiappori, P.-A. & Gollier, C., MIT Press.

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

H.M. TREASURY (2005). Economy charts and tables. *Pre-Budget Report*.

HOY, M. (2006). Risk classification and social welfare. *The Geneva Papers of Risk and Insurance: Issues and Practice*, **31**, 245–269.

HOY, M. & POLBORN, M. (2000). The value of genetic information in the life insurance market. *Journal of Public Economics*, **78**, 235–252.

HOY, M. & WITT, J. (2007). Welfare effects of banning genetic information in the life insurance market: The case of BRCA1/2 genes. *Journal of Risk and Insurance*, **74:3**, 523–546.

JONES, F. (2005). The effects of taxes and benefits on household income, 2004/05. *Office for National Statistics*.

MACDONALD, A.S. (2004), Genetics and Insurance, in *Encyclopaedia of Actuarial Science*, ed. Teugels, J. & Sundt, B., John Wiley, Chichister.

MACDONALD, A.S., PRITCHARD, D.J. & TAPADAR, P. (2006). The impact of multifactorial genetic disorders on critical illness insurance: A simulation study based on UK Biobank. *ASTIN Bulletin*, **36**, 311–346.

MEYER, D.J. & MEYER, J. (2005). Risk preferences in multi-period consumption models, the equity premium puzzle and habit formation utility. *Journal of Monetary Economics*, **52**, 1497–1515.

NORBERG, R. (1995). Differential equations for moments of present values in life insurance. *Insurance: Mathematics and Economics*, **17**, 171–180.

POLBORN, M., HOY, M. & SADANAND, A. (2006). Advantageous effects of regulatory adverse selection in the life insurance market. *Economic Journal*, **116**, 327–354.

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. *Journal of Risk and Insurance*, **66:4**, 531–550.

VILLENEUVE, B. (1999). Mandatory insurance and the intensity of adverse selection. *CREST/INSEE working paper*.

VILLENEUVE, B (2000), Life insurance, in *Handbook of Insurance*, ed. Dionne, G., Kluwer Academic Publishers.

VILLENEUVE, B. (2003). Mandatory pensions and the intensity of adverse selection in life insurance markets. *Journal of Risk and Insurance*, **70**, 527–548.

VON NEUMANN, J. & MORGENSTERN, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

WOODWARD, M. (1999). *Epidemiology: Study design and data analysis.* Chapman & Hall.

XIE, D. (2000). Power risk-aversion utility functions. *Annals of Economics and Finance*, **1**, 265–282.