

# MODELING THE IMPACT OF GENETICS ON INSURANCE\*

Angus S. Macdonald†

---

## ABSTRACT

The role of probabilistic models in the debate over genetics and insurance is discussed. A Markov model is used to show that, under quite extreme assumptions, adverse selection in life insurance ought to be controllable. The statistical problems of estimating small differences in mortality are discussed; these might limit the use of many genetic disorders as rating factors. The influence of the insurance industry on policy-making, especially through its support of research, is discussed. It is suggested that participating contracts are suitable and simple vehicles to carry the genetic risks in life insurance.

---

## 1. INTRODUCTION

### 1.1 Genetics, Society, and Insurance

It has long been known that some diseases run in families. Such knowledge is epidemiological in nature, because it is distilled from observations of populations without the underlying mechanism necessarily being revealed. The science of human genetics is concerned with understanding these mechanisms that turn inherited traits into observed diseases. In medicine, of course, it is entirely desirable to extend knowledge, and to remove uncertainty, but uncertainty is the bedrock of private, voluntary insurance. Therefore, genetics matters to insurers.

Genetics matters to society too, not just for medical reasons. Almost as soon as Mendelian inheritance was discovered, the idea of “improving the breed” was espoused, and eugenics emerged as a serious science. It is easy to forget just how much pioneering work in mathematical statistics, by Pearson, Fisher, and others, was motivated by an interest in eugenics. Today’s concerns come from quite the opposite direction: no person’s genetic inheritance is a priori better than any other person’s, with the corollary that discrimination on genetic grounds is deprecated.

Insurance matters to society. When essential services are funded by private insurance, access to insurance, at reasonable prices, is justifiably a political issue. Here the scientific basis of private, voluntary insurance intrudes, a little uncomfortably. If individuals can choose whether or not to buy insurance, and how much they will buy, underwriting is needed to ensure that a fair price is paid: otherwise adverse selection might occur. By definition this is discrimination; A must pay more than B for the same cover because A differs from B. This is not always acceptable to all concerned, but by and large a consensus has been reached that allows insurers enough of their scientific principles to make things work.

Politicians now grappling with these issues have to resolve the tension between the scientific principle of insurance and the powerful appeal to remove discrimination. Each side has adherents. Some argue that it is wrong to relinquish a scientific principle; that leads to a slippery slope (Pokorski 1997). Others do not agree that a scientific principle must have primacy over basic notions of justice (Moultrie and Thomas 1997; O’Neill 1997). To a pragmatist, the scientific basis of insurance is real, but often not precise in its application; the crux of the matter is perhaps this: how much adverse selection would it take to break the system? In the absence of quantitative answers, the issue is likely to be decided on its political merits alone. But, when science is in question, political merits do have limitations; in 1897 Indiana almost passed into law some mathematics that implied  $\pi = 9.2376$  (Beckman 1971).

---

\*This paper received a James C. H. Anderson Memorial Award (Honorable Mention) from the Actuarial Education and Research Fund.

†Angus S. Macdonald, F.F.A., Ph.D., is Professor in the Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, e-mail, A.S.Macdonald@ma.hw.ac.uk.

Quantitative answers require data and models to help us interpret the data.

## 1.2 This Paper

In this paper I survey the role of probabilistic models in tackling the questions raised by genetic testing and insurance. My main concern is to stress the importance of carrying out such modeling as can be done in advance of laws being passed and regulations being drawn up, so that we should not be faced with the actuarial equivalent of having to use  $\pi = 9.2376$  in our calculations.

I concentrate on life insurance, by way of example, but it is likely that greater problems will arise in disability and long-term-care insurance. The methods I describe are capable, in principle, of being adapted to other forms of insurance.

Section 2 describes briefly the relevant facts of underwriting and genetics and lists what a useful model must be able to do. Sections 3 and 4 outline some possible approaches, namely, Markov models and random future lifetime models. Section 5 gives some examples of costs of adverse selection based on simple Markov models, and Section 6 explores some of the properties of the Markov model used. In Section 7, I consider the statistical problem of detecting relatively small differences in mortality and the implications for underwriting. Section 8 describes some ways in which product design and marketing contribute to the problem and could contribute to solving the problem. Research issues, and the research role of the insurance industry, are discussed in Section 9. Conclusions are given in Section 10.

## 2. MODELS FOR INSURANCE AND GENETICS

### 2.1 Purpose

Briefly I consider the purpose served by a probabilistic model in insurance. It quantifies risk in terms of observable quantities (some of which may be rating factors) and sometimes in terms of unobservable quantities. These quantities define the heterogeneity of the population, observed or otherwise. The modeled outcomes can be used to determine

- a. The composition of underwriting classes
- b. The prices and reserves in respect of each underwriting class
- c. The effect of any remaining heterogeneity.

Conventional “good practice” really means the proper application of (a) and (b), with (c) acting as a

check. The central role of the model is obvious. In case the composition of the underwriting classes cannot be determined, the model is even more important, in helping to determine whether or not insurance along conventional lines is feasible at all. When restrictions on underwriting are being contemplated, therefore, we should expect that an appropriate effort will be made to construct, understand, and explain credible models.

At this stage in the development of genetic science, we can see that new risks are likely to emerge, but we have almost no data with which to attack the problem quantitatively. Even so, models can be of assistance: we can make assumptions that are plausibly extreme and try to bound the additional costs that might arise, and we can identify data that can usefully be collected in the future.

I cover these aspects in Sections 5, 6, and 9.

### 2.2 Insurance Underwriting

In life insurance, the salient fact is the broad extent of the ordinary rates (OR) class, namely, that group whose members can be offered insurance without any extra premium. In the U.K., about 95% are so accepted, another 4% are insurable at an increased rate of premium, and only about 1% are declined (Leigh 1990). Matters are similar in Europe and North America (Chuffart 1996; Pokorski 1996). The extent of the OR risk pool is such that it must include lives whose mortality is above average—Leigh (1990) said that it would typically extend to about 130–150% of aggregate mortality—and in many cases, a genetic predisposition would only affect an underwriting decision for a life near the boundary of an underwriting band.

Since different rates are offered to males and females, and to smokers and nonsmokers, the OR class is not quite as homogeneous as it might appear. However, the recent introduction of “preferred lives” underwriting has gone much further in dismantling the OR class into smaller risk groups. From the insurer’s point of view, genetic information is leading only to a continuation of an existing trend. So far, preferred lives underwriting is not as widespread in the U. K. as it is in the U.S.

My main interest lies in which of the broad classes of genetic disorder might appear in the various underwriting bands.

### 2.3 Human Genetics

The usual classification of genetic disorders is as follows:

- a. A monogenic disorder is an alteration to a single gene. These are rare, and often a sufferer will not survive to become economically active, so their presence in the insured population is even rarer. In some cases, the expected time of onset of symptoms, and of death, does fall in a relevant age group; Huntington's disease is an example. These conditions present additional risks of low incidence, but of large magnitude. Monogenic disorders are relatively well understood, since their genetic nature is often clear from family history. For the same reason, they are already allowed for in underwriting, so that the main impact of genetic tests might be to allow unaffected individuals to be offered OR.
- b. A chromosomal disorder is the alteration of genetic material on the larger scale of a chromosome, such as the presence of an extra chromosome in Down's syndrome. Symptoms are usually present throughout life, so that the question of using a genetic test to gain information about an asymptomatic individual does not arise.
- c. A somatic disorder is an alteration to genetic material that occurs after birth. It is localized (that is, not present in all the individual's genetic material), so it cannot usually be detected by a genetic test carried out on asymptomatic individuals.
- d. A multifactorial (polygenic) disorder is a combination of altered genes that, together with environmental and lifestyle factors, indicates a predisposition toward some disease, including common causes of death. These conditions present additional risks of high incidence, but often of low magnitude. Indeed, to the extent that an individual's lifestyle can be changed, given knowledge of the risk, mortality might even be improved by their detection.

The complex interactions of multifactorial disorders and the environment present the underwriter with quite difficult statistical problems; I consider these in Section 7.

Banning access to genetic test results (including family history) would lead to two distinct problems for life insurers. A very small number of lives with monogenic disorders, currently uninsurable, would be admitted to the pool, and a larger number of lives with multifactorial disorders would be rated more favorably than would otherwise be the case. However, given the extent of traditional underwriting bands, and the

relatively small extra risks presented by many multifactorial disorders, the latter lives might often be placed in the same underwriting band anyway.

We assume here that the definition of a genetic test is limited to examination of genetic material from asymptomatic lives and does not include diagnostic tests for lives with symptoms of disease that might be caused by a somatic disorder, for example. If genetic tests are defined more widely, insurers could be faced with much greater problems.

It is much harder even to hazard a guess about the impact of the various types of genetic disorder on other forms of insurance. Underwriting is typically more difficult to begin with, and the insured events are less well-defined and more capable of manipulation.

## 2.4 Requirements of a Model

In view of these issues, we can set out the properties needed if a model is to capture the main points of genetics, underwriting, and insurance:

- a. It must reflect heterogeneity of the population with respect to mortality, from genetic or other causes, and any other factors that affect the decision to buy insurance and to maintain it in force.
- b. It must represent the information available to applicants and insurers, including that which becomes available as a result of a genetic test.
- c. It must represent the incidence and outcomes of genetic testing.
- d. It must represent the decision to buy insurance, in light of the available information.
- e. It must represent the underwriting process, in light of the available information.
- f. It must embody realistic insurance payments.
- g. It should be probabilistic in nature, to allow uncertainty to be quantified. First moments suffice for naive pricing and reserving, but more information is desirable.
- h. It should be specified in terms of quantities that are, in principle at least, observable and capable of statistical estimation.

In the following sections I discuss two approaches: first, in Section 3, a Markov approach in the spirit of Hoem (1969, 1988) and Norberg (1995); second, in Section 4, a random future lifetime approach in the spirit of Bowers et al. (1986) and Gerber (1990). In both cases I use the idea of frailty to represent heterogeneity.

### 3. MARKOV MODELS

#### 3.1 Frailty

The ideas of frailties, and frailty distributions, will be useful. Vaupel, Manton, and Stallard (1979) introduced the term “frailty” to model heterogeneity in mortality studies. With each life a (random) frailty parameter  $Z$  is associated, and the force of mortality in respect of that life is a function of  $Z$ , say,  $\mu_{x+t}^{(Z)}$ . The overall mortality of a cohort of lives then depends on the sampling distribution of  $Z$ . A common assumption is that the frailty is a univariate random variable and acts multiplicatively:  $\mu_{x+t}^{(Z)} = Z\mu_{x+t}$ .

#### 3.2 Heterogeneity in a Markov Model

The chief features of the Markov approach are (a) a discrete frailty distribution, (b) an explicit representation of insurance-buying behavior by transitions between states, and (c) a complete specification in terms of transition intensities (which determines the statistical problem).

To begin, I suppose that heterogeneity can be described by a univariate frailty, with a discrete distribution on some set of positive real numbers:

$$z_1 < z_2 < \dots < z_{M-1} < z_M \quad (1)$$

with frailty distribution

$$P[Z = z_i] = p_i \quad (i = 1, 2, \dots, M)$$

and that it acts multiplicatively on a baseline force of mortality, which in these examples is given by

$$\mu_{x+t} = 0.00002072e^{0.103571(x+t)} \quad (2)$$

(which is broadly consistent with U.K. male assured lives data 1979–82). The frailty divides the population into  $M$  subpopulations. Within each subpopulation, the lifetime of an individual is defined by a sample path in a Markov model, such as that in Figure 1. In this example, a life in the  $i$ -th subpopulation starts in state  $i0$  at age  $x$ , uninsured and not yet having had a genetic test, and may progress between states as indicated. At outset, neither the life nor the insurer knows to which subpopulation the life belongs (that is,  $Z$  is unobserved at outset).

Such subpopulations can represent entire underwriting classes, or they can represent segments of underwriting classes. If applicants have better knowledge than insurers, an insurance class will be represented by a collection of subpopulations, within each of which insurance-buying behavior will differ.

This model was used by Macdonald (1997); the presentation there omitted technical discussion, some of which I supply here.

#### 3.3 An Alternative Formulation

The subpopulations defined by the univariate frailty  $Z$  are each homogeneous in respect of mortality, but not in respect of genetic status. Each might include lives with and without genetic disorders of any kind, which I suppose are distinguished by genetic tests. An alternative formulation, more convenient for some purposes, splits each of these subpopulations by genetic status, and possibly by the level of additional mortality conferred by genetic status. This amounts to using a bivariate frailty  $Z = (Z_1, Z_2)$ , in which  $Z_1$  represents overall mortality and  $Z_2$  represents genetic status, or additional mortality from genetic causes. Examples using both formulations are given in Sections 5 and 6.

#### 3.4 Genetic Testing

The extent of genetic testing in the  $i$ -th subpopulation is governed by the overall intensity  $\mu_{x+t}^{i01} + \mu_{x+t}^{i03}$  and the chance of a positive result is  $\mu_{x+t}^{i03}/(\mu_{x+t}^{i01} + \mu_{x+t}^{i03})$ . It is reasonable to suppose that the incidence of testing does not depend on  $i$ , but the chance of a positive test might depend on  $i$ .

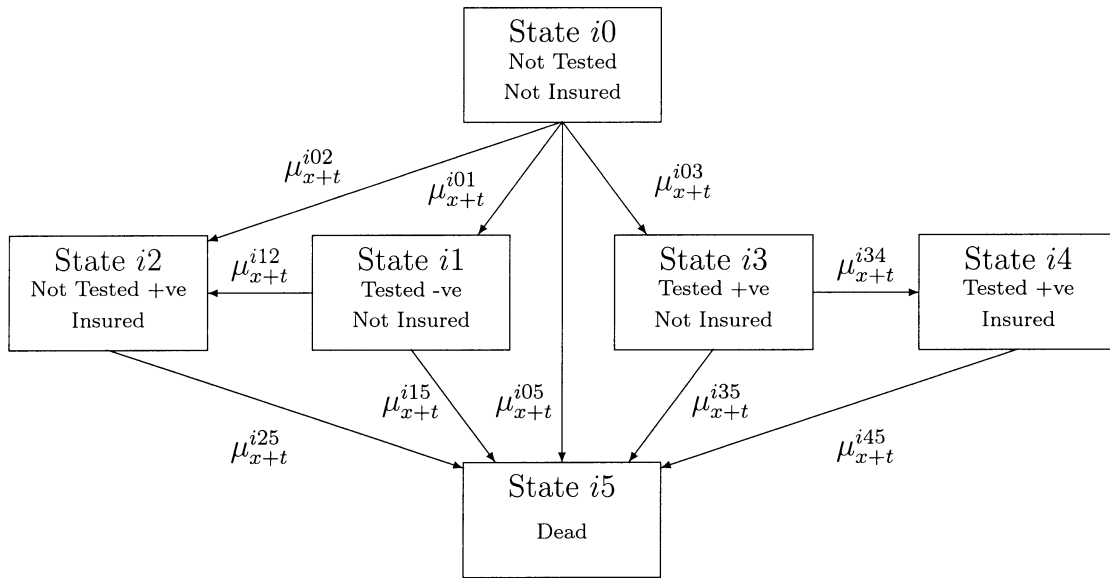
#### 3.5 Insurance-Buying Behavior

We suppose there is a “normal” transition intensity from uninsured to insured states. Within any underwriting band, and in the absence of genetic testing, this should not depend on  $i$ . Adverse selection could take two forms. If the  $i$ -th subpopulation suffers heavy mortality, it is represented by increasing the transition intensity  $\mu_{x+t}^{i34}$  or by increasing the sum assured payable on exit from state  $i4$ . If the  $i$ -th subpopulation enjoys light mortality, it is represented by decreasing the corresponding quantities.

#### 3.6 Underwriting

Underwriting is represented by varying the rate of premium payable, depending on the insured state occupied and on the ability (or inability) of the insurer to distinguish between subpopulations and between lives with and without genetic disorders. We can choose the frailty to model any system of underwriting, including preferred lives. By default, I assume that the OR class (which might include several subpopulations) is charged premiums calculated using the baseline mortality.

Figure 1  
A Markov Model for the  $i$ th of  $M$  Subpopulations



### 3.7 Insurance Payments

On exit from an insured state, a sum assured is payable. The basic sum assured is \$1, but this could be increased or decreased because of adverse selection. I denote the benefit payable on death while in state  $ik$  by  $b^{ik5}$ .

The traditional level premium, depending on age at entry, complicates the model, because lives entering an insured state at different ages would pay different premiums, so the payments (though not the transitions) would depend on more than the state currently occupied. This difficulty is simply avoided by adopting the system of paying for mortality risk as it arises. In the U.K., it is common practice under unit-linked contracts to deduct a monthly mortality charge from the fund, of amount  $q_x/12$  per unit sum assured (on some suitable basis). In my model, if a life is insured at age  $x + t$ , a premium is payable, at a rate per unit sum assured that is a multiple of  $\mu_{x+t}$  (the multiple depending on the underwriting).

Under this “current cost” system, no prospective mortality reserves are held if the mortality charges are calculated on the reserving assumptions. Technically, under the probability measure given by the reserving basis, the payment process formed by the discounted mortality cash flows (claims minus premiums) is a martingale. Compared with the traditional level premium system (under which reserves are accumulated out of premiums paid in excess of expected claims), this model will overstate the insurance losses caused

by adverse selection, which is desirable if we want to estimate upper bounds for such losses.

### 3.8 Results from the Model

The quantities of interest are the policy values, contingent upon the states occupied, if we extend the idea of “policy value” to those states in which a life has not yet bought insurance. Moments of these policy values provide a measure of costs, with and without adverse selection. Norberg (1995) showed that the moments satisfy a system of differential equations, which can be solved recursively (backwards) by standard numerical procedures, the boundary condition being that all policy values are zero at the end of the term. Norberg’s equations are given in the Appendix. Hesselager and Norberg (1996) have further shown how the distributions of present values could be computed, at least in principle. I give some examples of results from Markov models in Section 5.

## 4. AN ALTERNATIVE APPROACH: RANDOM LIFETIME MODELS

An intrinsic feature of the Markov models is that they represent heterogeneity as a discrete frailty, possibly regarded as an approximation to a continuously distributed frailty. We might ask: why not approach the problem using the well-known ideas of frailty models in the first instance? For example, if the frailty  $Z$  has

density  $f_{\varepsilon}(z)$ , the  $q$ -th moment about zero of a sum assured of  $s$  payable on death within  $n$  years is

$$\int_0^n \int_0^\infty (se^{-\delta t})^q \mu_{x+t}^{(Z)} e^{-\int_0^t \mu_{x+s}^{(Z)} ds} f_{\varepsilon}(z) dz dt, \quad (3)$$

which can easily be evaluated numerically. This would allow us to make use of existing studies of frailty distributions, such as Vaupel, Manton, and Stallard (1979), Hougaard (1984), and Vaupel (1988). It is also in the spirit of the random lifetime models found in many actuarial texts, for example, Bowers et al. (1986) and Gerber (1990). However, while frailty models perhaps offer a more flexible representation of heterogeneity, including some useful statistical tools, they do not offer any natural representation of insurance-buying behavior. For example, to calculate the moments of present values of the insurance loss, with premiums as in Section 3, we need to compute

$$\int \left( \int_r^t se^{-\delta(r+u)} \mu_{x+u} du - se^{-\delta t} \right)^q dF(r, t, s, \varepsilon), \quad (4)$$

where  $F(r, t, s, \varepsilon)$  is the joint distribution of  $r$ , the time at which insurance is purchased;  $t$ , the time at which insurance cover ceases by death or by expiry;  $s$ , the sum assured; and  $\varepsilon$ , the frailty. It is difficult to find any natural assumptions for  $F$ ; in fact, the most obvious approach might be to calculate the occupancy probabilities in respect of the various insured states in a Markov model such as we have used. Then the moments given by Norberg's equations can be seen to be numerical estimates of the integrals in Equation (4).

A further complication is that a univariate frailty subsumes, into a single variable  $Z$ , all factors that influence mortality. We soon find it necessary to distinguish the separate contribution of a genetic predisposition, so we might consider bivariate frailties  $Z = (Z_1, Z_2)$  in the first instance, along with some assumption about the action of the frailty on the force of mortality. For example, models with frailties acting multiplicatively such as

$$\mu_{x+t}^{\varepsilon} = \beta_1 \varepsilon_1 \beta_2 \varepsilon_2 \mu_{x+t} \quad (5)$$

$$\mu_{x+t}^{\varepsilon} = (\beta_1 \varepsilon_1 + \beta_2 \varepsilon_2) \mu_{x+t} \quad (6)$$

for some constants  $\beta_1$  and  $\beta_2$  are quite natural (the former can be interpreted as a Cox model with unobserved covariates). However, we have lost the link with existing statistical work on frailties, which is largely in a univariate framework.

Extending this approach to other forms of insurance, especially disability insurance with repeated

transitions, presents formidable numerical difficulties, because the calculation of moments requires the approximate evaluation of high- or infinite-dimensional multiple integrals.

For these reasons, random lifetime models are less promising than they might appear to be, and provided we have sufficient computational power to solve Norberg's equations with a reasonably large number of subgroups, the Markov approach is preferred.

## 5. EXAMPLES

### 5.1 Data

In this section I give some examples of the analyses that can be carried out using the Markov model in Figure 1. The first obstacle we encounter is the lack of data from which to estimate transition intensities. We can obtain just a few pointers (the following are based on the U.K.):

- a. Following Section 2.3, as a first approximation, we can model separately the impact of multifactorial disorders on the current OR group, and more severe disorders on the current rated-up and declined groups.
- b. About 95% of applicants have mortality of up to 150% of the average. The proportion of the whole population ought to be lower, since applicants are likely to self-select, but at least we can roughly identify the OR group.
- c. Numbers of life insurance policies bought each year are available.
- d. Statistics are available in respect of a few monogenic disorders (Harper 1997), but few useful data are available in respect of multifactorial disorders.

In the circumstances, all that we try to do is to make assumptions that tend to overstate the additional costs of adverse selection in life insurance, while pointing out the need for research if better estimates should be required (see Section 9).

### 5.2 Modeling the Costs of Adverse Selection

Macdonald (1997) applied the Markov model of Figure 1 to several cases:

- a. The effect of multifactorial disorders on an OR class with two subpopulations (low and high mortality)
- b. The effect of more severe genetic disorders on a combined OR and rated-up class with two subpopulations

- c. The effect of severe late-onset monogenic disorders on a combined OR and rated-up class with two subpopulations
- d. The effect of all genetic disorders on a combined insured population with three subpopulations.

Here I give some examples from these results; in Section 6 I explore further the model of the OR class ([a] above). To give a baseline, I begin with a single homogeneous population, subject to the mortality of Equation (2), with no genetic testing. I suppose that the transition intensity from uninsured into insured states (the “force of insurance”) is a constant 0.05. (Obviously, this assumption could be refined, but it is probably too low at young ages and too high at old ages, so the costs of adverse selection will err in the right direction.) The expected present values of a sum assured of \$1 at force of interest  $\delta = 0.05$ , for various ages and terms, are shown in Table 1, which is referred to as the “baseline model.” The figures in respect of the insured state are those that the actuary would use in the traditional equation of value; they are the costs in respect of a life chosen at random from those who have just bought insurance. The figures in respect of the uninsured state represent the costs in respect of a life chosen at random from the uninsured population; I measure the costs of adverse selection as a proportion of these expected present values.

Figure 2 shows a simple model of the OR class with two subpopulations: a low-mortality group with  $Z = 0.75$ , and a high-mortality group with  $Z = 1.25$ . I retain the assumption that the normal force of insurance is 0.05 and represent adverse selection by increasing  $\mu_{x+t}^{234}$  (indicated in the figure) or the sum assured paid on death while in state 24, denoted  $b^{245}$ .

We calculate moments of insurance costs using Norberg’s equations (Norberg 1995), allowing for the unknown frailty. That is, at outset we know that a life is in state 10 or state 20, but we do not know which.

A feature of this model is that the subpopulations are distinguished only by their overall mortality; additional mortality related to genetic disorders is not

accounted for separately. Instead, I suppose that there is a higher incidence of genetic disorders in the high-mortality subpopulation (see the comments on univariate frailties in Sections 3 and 4). In fact, we make the extreme assumption that genetic disorders are completely absent from the low-mortality subpopulation.

The incidence and outcome of genetic testing in respect of multifactorial disorders is unknown. I illustrate two levels, as shown in Table 2. In both cases, 20% of tests have a positive (that is, adverse) outcome.

The rate of adverse selection is represented by the force of insurance  $\mu_{x+t}^{234}$ . Compared with the “normal” intensity of 0.05  $\mu_{x+t}^{234} = 0.25$  is high and  $\mu_{x+t}^{234} = 1.0$  is extremely high; the latter means that a life who has tested positive is practically certain to buy insurance within a few years. Table 3 shows the additional costs of adverse selection in this model, with these levels of adverse selection, with low and high levels of genetic testing, and with sums assured taken out by “adverse selectors” equal to one, two, or four times the average. To be precise, expected present values per sum assured of \$1 were first calculated with no adverse selection (these were almost the same as in Table 1); then expected present values were calculated in the presence of adverse selection; and the excess costs were expressed as percentages of the former.

I omit the details of the other cases considered by Macdonald (1997). Overall, the conclusions were that (1) in terms of order of magnitude, additional costs arising from adverse selection were more likely to be 10% than 100% (inevitably this has been quoted incorrectly in the press as “not more than 10%”), and (2) above-average sums assured was the most expensive aspect of adverse selection.

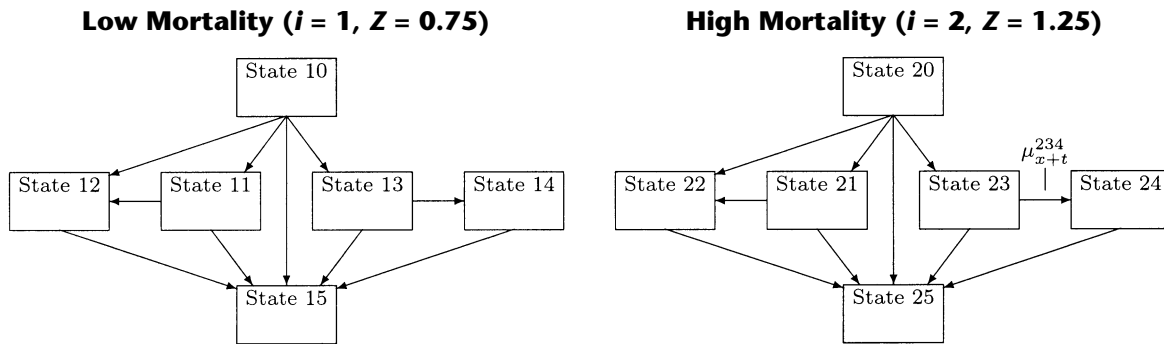
These conclusions broadly support restricting the sums assured that might be obtained without having to disclose genetic information to amounts not above the average. In the U.K., the Association of British Insurers currently does not require the results of any

**Table 1**  
**Baseline Model: Expected p.v. at  $\delta = 0.05$  of \$1 Death Benefit, Baseline Mortality, No Genetic Testing**

State at Outset	Age 30			Age 40		Age 50
	10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
Insured	0.00611	0.01638	0.03322	0.01708	0.04507	0.04722
Uninsured	0.00141	0.00688	0.01894	0.00392	0.01881	0.01078

Source: Macdonald 1997.

Figure 2  
A Markov Model of the OR Class with Two Subpopulations



genetic test to be disclosed if the sum assured does not exceed £100,000, and if the policy is being taken out in connection with a mortgage.

**5.3 Extensions of the Model**

It is not difficult to modify the basic model in Figure 1 to represent different contracts, or to bring in more features of life insurance. For example, Pritchard (1997) studied the following hypotheses:

- a. Genetic testing takes place from birth, so there is a pool of tested lives when insurance-buying ages are reached (20 onwards)
- b. Lives who buy insurance before being tested may be tested later, and if the result is positive, they may become adverse selectors and increase their sums assured
- c. Lives in the low-mortality subpopulation who test negative have a lower force of insurance (0.01, 0.005, or 0.0025), but no tendency to opt for lower sums assured
- d. There is selective lapsing from the insured states, depending both on the outcomes of genetic tests and on the sums assured of adverse selectors.

With lives who test negative in the low-mortality subpopulation either 5 or 20 times more likely to lapse their policies (called low and high selective lapsing respectively), and with lives who test positive in the high-mortality subpopulation less likely to lapse

their policies (5, 10, or 20 times less likely, for lives with sums assured of \$1, \$2, or \$4, respectively) the expected costs of adverse selection are given in Table 4. Since it is assumed that 80% of tests are negative, selective “absentment” ([c] above) and selective lapsing are expensive (particularly when lapse rates are high to begin with), but once more the greatest costs are associated with adverse selectors opting for higher sums assured.

Tan (1997) applied the model to purchasers of immediate annuities. Assuming that annuities are bought between ages 50 and 70, by which time a pool of tested individuals exists, she obtained additional costs of adverse selection of up to 40%. However, the assumptions leading to such high figures were sufficiently extreme (for example, a high level of genetic testing, with negative outcomes in 97% of cases, and lives in the low-mortality subpopulation who test negative up to 20 times more likely to buy an annuity, or up to 6 times the average amount) that she concluded that 10% was once again a reasonable order of magnitude for additional costs.

**6. EXPLORING THE MODEL OF THE OR CLASS**

The model described in Section 5.2 was based on some crude approximations. While many of these cannot easily be refined with the available data, we can at least explore the sensitivities of the model to the assumptions. Further, it is not only the expected costs of adverse selection that are of interest, higher moments are needed, for example, in support of risk-based capital margins. In this section I consider the development of the model of the OR class.

Some of the key simplifications in Macdonald (1997) are as follows:

Table 2  
Levels of Genetic Testing in High-Mortality Subpopulation

Level of Testing	$\mu_{x+t}^{201}$	$\mu_{x+t}^{203}$
Low	0.04	0.01
High	0.20	0.05

**Table 3**  
**Ordinary Rates Class: Mean p.v. of Loss, as a Percentage of Baseline Costs, with Two Levels of Genetic Testing and Three Levels of Sum Assured  $b^{245}$**

Level of Testing	$\mu_{x+t}^{234}$	$b^{245}$	Age 30			Age 40		Age 50
			10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
Low	0.25	1	0.7%	0.7%	0.4%	0.8%	0.5%	0.6%
		2	2.1	2.2	1.8	1.8	2.0	1.7
		4	4.3	5.1	4.8	4.1	4.8	4.0
	1.00	1	1.4	1.0	0.6	1.3	0.9	1.2
		2	3.6	2.8	2.2	3.3	2.6	3.1
		4	7.1	6.4	5.5	6.9	6.2	6.7
High	0.25	1	2.9	1.9	1.1	2.6	1.7	2.3
		2	6.4	5.4	4.1	6.4	5.2	6.0
		4	13.6	12.7	10.4	13.8	12.4	13.6
	1.00	1	4.3	2.5	1.3	4.3	2.3	4.3
		2	10.0	6.7	4.7	10.2	6.5	9.9
		4	21.4	15.1	11.4	21.7	14.9	21.4

Source: Macdonald 1997.

- a. A univariate frailty with two values was used ( $Z = 0.75$  or  $1.25$ ). I regard this as a discretization of a uniformly distributed continuous frailty on  $[0.5, 1.5]$  (motivated by Leigh 1990). An obvious refinement is then to use a finer discretization.
- b. It was assumed that genetic predispositions were absent from the low-mortality subpopulation. If we discretize more finely, this seems less reasonable.
- c. With a univariate frailty, the only link between genetic predispositions and mortality is a higher incidence of predispositions in subpopulations with higher frailties. With a bivariate frailty, we can assume explicitly that a genetic predisposition confers some additional mortality.

In this section I consider the effect of these assumptions on the moments of the insurance costs with and without adverse selection. I assume that the incidence of adverse selection, when it is present, is always severe: the transition intensity is  $\mu_{x+t}^{i34} = 1.0$  in respect of affected lives.

**6.1 Discretizing a Uniform Frailty**

Table 5 shows the first three moments (mean, variance, and skewness) of the claims costs and the insured loss, for various ages and terms, in the absence of adverse selection. Note that:

**Table 4**  
**Ordinary Rates Class: Mean p.v. of Loss, as a Percentage of Baseline Costs, with Selective Lapsing**

Baseline Lapse Rate	Selective Lapsing	S.A. of Adverse Selectors	Adverse Selection/Genetic Testing in Group 2			
			Low/Low	Low/High	High/Low	High/High
0.125	Low	1	14.0%	16.2%	13.9%	16.0%
		2	26.9	31.8	27.2	31.9
		4	54.2	64.7	55.3	65.7
0.125	High	1	15.6	16.9	14.5	16.7
		2	27.6	32.5	27.8	32.6
		4	54.8	65.4	56.0	66.4
0.020	Low	1	6.3	7.3	6.3	7.4
		2	9.3	10.8	9.4	11.0
		4	15.3	17.7	15.6	18.1
0.020	High	1	7.1	8.1	7.2	8.2
		2	10.1	11.6	10.3	11.7
		4	16.1	18.5	16.5	18.9

Source: Pritchard 1997.

**Table 5**  
**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Claims Costs and of Loss, per Unit Sum Assured, with Two Subpopulations and No Adverse Selection**

Present Value of	$q$	Age 30			Age 40		Age 50
		10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
Claims	1	0.001405	0.006874	0.018888	0.003920	0.018762	0.010759
	2	0.00101	0.00351	0.00690	0.00281	0.00963	0.00771
	3	0.0007	0.0019	0.0029	0.0020	0.0052	0.0056
Loss	1	0.000000	-0.000007	-0.000053	-0.000003	-0.000054	-0.000026
	2	0.00101	0.00351	0.00690	0.00281	0.00963	0.00772
	3	0.0007	0.0018	0.0026	0.0020	0.0048	0.0052

- a. I show fewer significant figures in respect of the higher moments, because Norberg's equations are solved recursively, and for a given step-size in the numerical algorithm, higher moments have higher errors. The significant figures shown are sufficient for my purposes.
- b. The variances of the claims costs and of the losses are practically the same. This reflects the well-known fact that under basic regular premium life insurance contracts, the variance is dominated by the rare, but large, payments made on death; the premiums contribute a more nearly deterministic series of cashflows.
- c. The expected claims costs are not quite the same as those in Table 1, because that table was based on a single homogeneous population, while Table 5 is based on two subpopulations. They are very close, however, because mortality rates are so small at ages 30–60 that the expected costs are almost linear in the frailty. This would not be the case for annuity business (Tan 1997).
- d. The expected losses are not quite zero, because all lives are charged the same aggregate mortality premiums, rather than lives in each subpopulation being charged their "true" premiums, but again the expected losses are nearly linear in the frailty at these ages, so the discrepancy is very small.

Finer discretizations are obtained by using four subpopulations, with 62.5%, 87.5%, 112.5%, and 137.5% of the average mortality; eight subpopulations, with 56.25%, 68.75%, . . . , 143.75% of the average mortality, and so on. Table 6 shows the first three moments of the present value of the loss, in the absence of adverse selection, for a life aged 30 at outset, and a term of 30 years, with 2, 4, 8, 16, or 32 subpopulations. This shows that (1) the crude discretization in the model with two subpopulations overstates the expected loss, although it is still almost zero, (2) the

discretization has very little effect on the higher moment, and (3) further subdivision would not be worthwhile in this example.

Table 7 shows the moments of the present value of the loss for the same ages and terms as in Table 5, with no adverse selection, and 32 subpopulations.

It is worth commenting on the assumption of a uniform frailty, which was reasonably sensible with only two subpopulations, but that I might have tried to refine at the same time as refining the discretization. For example, motivated by Vaupel, Manton, and Stallard (1979), I might have used a truncated gamma distribution. I have not done so, because there is little evidence to support any particular form of frailty in respect of the OR class, so I preferred to retain an assumption that, arguably, overstates the incidence of higher frailties and understates the incidence of "average" frailties. This is just one of many examples of assumptions that must await better knowledge of heterogeneity.

## 6.2 The Incidence of Genetic Predispositions

In Section 5, with two subpopulations, I assumed (1) that 20% of genetic tests in the high-mortality subpopulation had a positive result, and (2) that all tests were negative in the low-mortality subpopulation.

With more than two subpopulations, it is not realistic to suppose that the incidence of genetic predispositions depends solely on whether  $Z > 1$  or  $Z < 1$ . However, as a first step, I modify only supposition (2) above and suppose that 4% of genetic tests in the low-mortality subpopulation are positive.

It is sensible to assume that the incidence of genetic testing does not depend on the frailty (since the frailty is unknown to the individual); in terms of Section 5,  $\mu_{x+t}^{i01} + \mu_{x+t}^{i03}$  is the same for all  $i$  (0.25 in these examples). For the same reason, it also seems sensible

Table 6

**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Claims and of Loss, per Unit Sum Assured, for a Life Aged 30 at Outset and Term 30 Years, with No Adverse Selection**

Present Value of	$q$	Number of Subpopulations				
		2	4	8	16	32
Claims	1	0.018888	0.018876	0.018872	0.018872	0.018871
	2	0.00690	0.00689	0.00689	0.00689	0.00689
	3	0.0029	0.0029	0.0029	0.0029	0.0029
Loss	1	-0.000053	-0.000067	-0.000070	-0.000071	-0.000071
	2	0.00690	0.00690	0.00689	0.00689	0.00689
	3	0.0026	0.0026	0.0026	0.0026	0.0026

to suppose that insurance-buying behavior, following a genetic test, is the same in all subpopulations. Table 8 shows the moments of the insured losses, with sums assured of \$1, \$2, or \$4 in respect of adverse selectors.

Compared with Table 7, the interesting point is not the expected losses (which of course are no longer negligible) but the higher moments. If all sums assured are \$1, these are increased fairly modestly by the adverse selection, but if adverse selectors opt for higher sums assured, they are greatly increased. I emphasize that I have tried to choose extreme assumptions, but the following conclusion seems plausible:

- a. If insurers are exposed to adverse selection in respect of the propensity to buy insurance alone, there should be little effect on risk-based capital needs.
- b. If, however, adverse selection in the form of higher sums assured is also possible, higher risk-based capital margins might be indicated, even if the premiums charged are adequate.

In Table 8 I refined the discretization of the frailty, but not the way that the incidence of genetic predispositions might depend on the frailty. I now do that. I have absolutely no basis in fact for any particular assumption, but in the spirit of seeking to bound the expected costs, I amend the assumption of Section 5 as follows:

- a. Lives with  $Z = 0.5$  have no genetic disorders
- b. The “average” probability that a genetic test is positive for  $Z < 1$  is 0.04
- c. The “average” probability that a genetic test is positive for  $Z > 1$  is 0.20
- d. The probability that a genetic test is positive is a piecewise linear function of the frailty.

This results in probabilities that a genetic test is positive as shown in Table 9. Considering that extra mortality of +50% is by no means high, it is possible that these assumptions are severe.

Table 10 shows the mean, variance, and skewness of the losses with sums assured of \$1, \$2, or \$4 in respect of adverse selectors.

Comparing these with Table 8, we see that the expected losses have increased by up to 30%. Clearly, the crude use of an average incidence of genetic disorders of  $Z > 1$  is not negligible if that incidence is in fact more strongly related to the frailty. However, the second and third moments are only slightly different, so our conclusion remains the same as above.

**6.3 A Bivariate Frailty**

As mentioned previously, genetic predispositions do not confer any additional mortality within any subpopulation; their influence is solely through their incidence. This is a consequence of representing heterogeneity by a univariate frailty. If, instead, we used a

Table 7

**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Loss, per Unit Sum Assured, with No Adverse Selection, in a Markov Model with 32 Subpopulations**

$q$	Age 30			Age 40		Age 50
	10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
1	-0.000001	-0.000009	-0.000071	-0.000004	-0.000072	-0.000034
2	0.00100	0.00351	0.00689	0.00281	0.00962	0.00771
3	0.0007	0.0018	0.0026	0.0020	0.0048	0.0052

**Table 8**  
**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Loss,**  
**for Sums Assured  $b^{i45} = \$1, \$2, \text{ or } \$4$  in Respect of Adverse Selectors,**  
**in a Markov Model with 32 Subgroups**

$b^{i45}$	$q$	Age 30			Age 40		Age 50
		10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
1	1	0.000050	0.000133	0.000163	0.000137	0.000314	0.000350
	2	0.00126	0.00407	0.00764	0.00353	0.01116	0.00968
	3	0.0009	0.0022	0.0030	0.0025	0.0056	0.0066
2	1	0.000114	0.000363	0.000666	0.000316	0.000938	0.000836
	2	0.00224	0.00668	0.01185	0.00624	0.01823	0.01707
	3	0.0026	0.0055	0.0072	0.0070	0.0143	0.0184
4	1	0.000243	0.000824	0.001671	0.000673	0.002186	0.001807
	2	0.00613	0.01708	0.02871	0.01708	0.04654	0.04665
	3	0.0157	0.0325	0.0410	0.0432	0.0839	0.1131

bivariate frailty (see Sections 3 and 4) we could (1) split each subpopulation in Figure 1 into two, one containing those lives possessing a genetic predisposition, and one containing those lives free of a genetic predisposition, and (2) model explicitly the additional mortality conferred by a genetic predisposition. To some extent this is just bookkeeping. It matters little whether a life whose overall frailty is  $Z = 1.1$ , say, and who has a predisposition that confers additional mortality of 10%, is regarded as having a univariate frailty with value 1.1, or a bivariate frailty with “basic” component 1.0 and “genetic” component 0.1.

As an illustration, suppose that a genetic predisposition confers additional mortality of 10%, otherwise making the same assumptions as in the previous section. The figure of 10% is not based on sound data but might not be unreasonable.

If we separate out a genetic component of mortality, it is questionable whether we should continue to suppose that the incidence of positive tests increases with the “basic” component of the frailty. I do so here because it is on the conservative side, and because there is no real scientific basis for the assumption of a uniform 10% addition to mortality from genetic causes.

**Table 9**  
**Probability That a Genetic Test Has a Positive Result, as a Piecewise Linear Function of the Frailty in the OR Class**

Frailty $Z$	0.50	0.75	1.00	1.25	1.50
P [positive result]	0.00	0.04	0.08	0.20	0.32

Table 11 shows moments of the present values of claims costs in the absence of genetic testing. These are only slightly higher than the corresponding figures in Table 5, so with no genetic tests or adverse selection it makes little difference what type of frailty we use.

Table 12 shows the costs of adverse selection, corresponding to Table 10. The effect of the additional mortality is considerable. If the sums assured of adverse selectors are not above average, the mean additional costs are roughly doubled. If adverse selectors opt for higher sums assured, the increase is proportionately much less, because higher sums assured is already the costliest part of adverse selection. Higher moments are almost the same as in Table 10, however.

To put the results of the this section into perspective, I recapitulate the conservative elements of the assumptions culminating in Table 12:

- A high rate of adverse selection ( $\mu_{x+t}^{i34} = 1.0$  in respect of affected lives)
- A discretized uniform frailty (32 subgroups), possible overstating the incidence of higher frailties in the OR class
- An incidence of genetic predispositions that increases with the frailty, as in Table 9
- Additional mortality of 10% for lives with a genetic predisposition.

I have only considered the OR class here, and I have not included the other facet of adverse selection, namely, any tendency for healthier lives to abstain from buying insurance [see Pritchard (1997)]. Even so, when the additional expected losses in Table 12 are expressed as a percentage of the baseline claims costs in Table 11—see Table 13—the results bear out

**Table 10**  
**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Loss, with Sums Assured  $b^{i45} = \$1, \$2, \text{ or } \$4$  in Respect of Adverse Selectors, in a Markov Model with 32 Subgroups in Which the Chance of a Positive Test Depends Linearly on the Frailty**

$b^{i45}$	$q$	Age 30			Age 40		Age 50
		10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
1	1	0.000067	0.000180	0.000242	0.000184	0.000444	0.000479
	2	0.00128	0.00410	0.00767	0.00356	0.01122	0.00976
	3	0.0009	0.0022	0.0030	0.0025	0.0057	0.0066
2	1	0.000153	0.000487	0.000913	0.000422	0.001277	0.001127
	2	0.00230	0.00682	0.01207	0.00640	0.01860	0.01748
	3	0.0026	0.0057	0.0074	0.0073	0.0148	0.0190
4	1	0.000324	0.001102	0.002255	0.000898	0.002943	0.002424
	2	0.00637	0.01770	0.02964	0.01774	0.04811	0.04835
	3	0.0164	0.0339	0.0428	0.0451	0.0876	0.1181

the previous conclusions of Macdonald (1997). Provided the sums assured taken out by adverse selectors can be controlled, 10% seems to be a plausible order of magnitude of the additional losses.

It is worth repeating the warning that it would be extremely unwise to suppose that these conclusions need have any bearing on other types of insurance business.

### 7. STATISTICAL ISSUES

For genetic information to be translated into insurance costs, mortality studies must be carried out: suitable data must be collected and analyzed. Here I consider some problems that might arise, in the context of applying the results to underwriting or reserving.

I emphasize that insurers have a legitimate interest in genetic information, whether or not it is available to underwriters. If it is available, its purpose is to help in pricing and reserving. If it is not available, insurers

still need to analyze the risks in their portfolios, including those relating to heterogeneity. Genetic information might therefore be important in reserving, even if it cannot be used for pricing.

#### 7.1 Data Collection

The kind of study needed is simple in concept: observe a group of lives with a given gene or set of genes, and record the illnesses to which they succumb and the ages at which they die. The second part of this prescription is the basic stuff of survival analysis, and the key question is whether or not sufficient data are available to obtain reliable results. The first part—data collection—might prove to be difficult.

Insurers can only analyze the effect of rating factors if they obtain the relevant information from the applicant. If genetic information need not be disclosed, analysis is impossible. Then the only source of data would be medical studies, as at present, so we ought to consider why such studies are undertaken.

**Table 11**  
**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Claims Costs and of Loss, per Unit Sum Assured, with No Adverse Selection, if a Genetic Predisposition Imparts Additional Mortality of 10%**

Present Value of	$q$	Age 30			Age 40		Age 50
		10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
Claims	1	0.001421	0.006952	0.019083	0.003965	0.018956	0.010872
	2	0.00102	0.00355	0.00697	0.00284	0.00973	0.00780
	3	0.0007	0.0019	0.0029	0.0021	0.0053	0.0057
Loss	1	0.000017	0.000080	0.000206	0.000046	0.000204	0.000119
	2	0.00102	0.00355	0.00696	0.00284	0.00972	0.00779
	3	0.0007	0.0019	0.0027	0.0020	0.0048	0.0053

**Table 12**  
**Ordinary Rates Class: Mean, Variance, and Skewness ( $q = 1, 2, 3$ ) of p.v. of Loss,**  
**with Sums Assured  $b^{i45} = \$1, \$2, \text{ or } \$4$  in Respect of Adverse Selectors,**  
**if a Genetic Predisposition Confers Additional Mortality of 10%**

$b^{i45}$	$q$	Age 30			Age 40		Age 50
		10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yrs
1	1	0.000114	0.000347	0.000605	0.000315	0.000896	0.000836
	2	0.00131	0.00419	0.00780	0.00365	0.01144	0.01000
	3	0.0009	0.0022	0.0031	0.0026	0.0058	0.0068
2	1	0.000238	0.000793	0.001577	0.000661	0.002102	0.001775
	2	0.00241	0.00712	0.01252	0.00671	0.01936	0.01829
	3	0.0028	0.0060	0.0079	0.0077	0.0156	0.0202
4	1	0.000487	0.001684	0.003521	0.001351	0.004516	0.003654
	2	0.00682	0.01885	0.03138	0.01895	0.05102	0.05147
	3	0.0177	0.0365	0.0461	0.0486	0.0942	0.1270

The insurer is interested in epidemiological studies, yielding information on the incidence of morbidity and mortality. In genetic science, however, such studies are not of interest in themselves, but only as a means to identify a genetic factor in some disease process. Once the existence of a genetic link has been established, the problem is to find the gene(s) involved and to unravel the biochemical mechanism that produces the effect. Epidemiological studies do not contribute much to these stages.

The insurer wants to quantify an effect with reasonable reliability, if it is to be used in underwriting. More data are needed to do this than simply to establish the existence of an effect. Therefore, epidemiological studies sufficient to answer the geneticist's questions often will not answer the insurer's questions.

**7.2 The Reliability of Underwriting**

In the U.K. the Disability Discrimination Act 1995 (Services and Premises Regulations 1996) allows premiums to take account of disability, provided the

differences are (1) based on relevant information, (2) based on information upon which it is reasonable to rely, and (3) reasonable having regard to any other relevant factors.

It seems reasonable to assume that if insurers are allowed to use genetic information in underwriting, some similar strictures will apply. Then (3) might be especially significant in the case of multifactorial disorders, in view of the contribution of environment and lifestyle to the overall risk. It is perhaps unlikely that sufficient data will accrue from medical and epidemiological studies to allow reliable ratings to be derived.

I illustrate the statistical difficulties with a simplified example. My simplifying assumptions all tend to reduce the uncertainty in the analysis; any real study would be more difficult. The technical details are these: I simulate a number of random future lifetimes from two populations: a control population with constant hazard rate 0.04 and a second population with constant hazard rate 0.05. The statistical problem is to quantify the extra mortality of the second population; we know that it is 25%, but the statistician does not.

This experiment allows us to observe all the lifetimes (no censoring) and ignores all other factors (no covariates). Were these features present, much more data would be needed to attain any given degree of precision. Many studies of multifactorial disorders might be based on quite small proportions of the whole population, lifetimes will be quite long (so censored observations might be the rule), and there will often be many covariates.

The first line of Table 14 shows approximate 95% confidence intervals for the additional mortality (true value 25%), for sample sizes of 100, 1,000, and 10,000

**Table 13**

**Ordinary Rates Class: Mean p.v. of Loss, with Sums Assured  $b^{i45} = \$1, \$2, \text{ or } \$4$  in Respect of Adverse Selectors, as a Percentage of Baseline Costs, with All Worst-Case Assumptions**

$b^{i45}$	Age 30			Age 40		Age 50
	10 Yr	20 Yr	30 Yr	10 Yr	20 Yr	10 Yr
1	6.9%	3.8%	2.1%	6.8%	3.6%	6.6%
2	15.6	10.3	7.2	15.5	10.0	15.2
4	33.1	23.1	17.4	32.9	22.7	32.5

**Table 14**  
**95% Confidence Intervals for Additional Mortality from Three Random Samples**

Sample Number	Number in Each Population		
	100	1,000	10,000
1	12-95%	19-41%	21-28%
2	-24-33	17-39	21-28
3	8-89	11-33	22-29

in each population (see Collett 1994, Section 4.4 for details).

Taking the first interval as an example, with 100 lives in each population, we can say only that, with 95% confidence, the second population has extra mortality of between 12% and 95%. Increasing the sample sizes narrows the range considerably, but I remarked above how difficult this might be in practice. The question for the underwriter, and perhaps for the lawyer, is whether or not a premium rating based on this information is defensible under any given criteria.

However, we are not yet finished with the effects of uncertainty. The confidence interval above (12-95%) is itself based upon a random sample. Were we to repeat the experiment with a different sample, we would get a different result. To illustrate, Table 14 shows the 95% confidence intervals based on three independent samples. Especially with the smaller samples, there is a good deal of uncertainty even about how uncertain we ought to be. Naturally, this does not help to defend the objectivity of any premium rating.

Table 15 shows a similar range of confidence intervals, with a control population ten times larger than the second population. If the difference in mortality is small, even very large amounts of unusually “clean” data leave considerable doubt about the answer to the question “how big a difference is there?” Censoring

**Table 15**  
**95% Confidence Intervals for Additional Mortality from Three Random Samples, Control Population 10 Times Larger than Second Population**

Sample	Control Population/ Second Population	
	1,000/100	10,000/1,000
1	5-58%	22-39%
2	0-50	15-30
3	5-58	14-30

and covariates will make matters much worse in practice.

It is in this light that the likelihood of insurers being able to make their underwriting more precise, on the basis of genetic tests for multifactorial disorders, should be considered.

## 8. PRODUCT DESIGN AND INSURANCE REGULATION

The examples in Sections 5 and 6 illustrate losses under contracts which are mispriced because of asymmetric information. The insurance industry’s first and natural response is to try to remedy the asymmetry, but that effort might fail in the face of (genuine) public concerns about access to services that are perceived as necessities, and provided by the insurance industry.

In the U.K. the insurance industry is probably about to be offered a larger role in providing many services hitherto provided by the state. State provision has generally been at a basic level for all, so insurance companies, in the past, have mostly marketed appealing alternatives to affluent customers (including companies providing for employees through group schemes). The market has been fairly free from restrictions on pricing and underwriting. It is not obvious that this model is well suited for the new role on offer.

It is not up to insurers to argue for or against any particular forms of risk pooling, or cross-subsidy, provided only that the implications of any given scheme have been properly understood by policymakers. Therefore, insurers should not argue against restrictions on underwriting simply because they would increase cross-subsidies. That is more the pleading of an industry lobby than an application of scientific principle, and the more noisily it is advanced, the more any genuinely scientific argument risks being lost in the noise. I contend that, if insurers wish to earn profits from the social role that is increasingly being offered, they have to be able to do so in ways acceptable to the customer (that is, government) or else show convincingly why scientific principle stands in the way. We can ask: what is the likely response of policy-makers to conclusions such as those in Sections 5 and 6? I offer my guess: it will be one of surprise and hardening attitude, if the insurance industry cannot solve a problem of such relatively modest proportions.

If we strip the problem down to fundamentals, the crucial points are these:

- a. The risks being underwritten are foreseeable in qualitative terms, but are not easily quantified in advance
- b. Adverse selection does pose a risk, but one that is much reduced by sum assured limits
- c. There are only three parties who could pay any additional costs: taxpayers, shareholders, or policyholders.

Consideration of each of the three parties in (c) suggests a slightly different mechanism, though all have in common that policyholders will pay more through premiums or taxes.

If premiums are not immediately increased to a level sufficient to absorb the risk, either shareholders will bear the risk (in which case their risk/return requirement should eventually force premiums up) or government will have to act as reinsurer of last resort, which will push the cost onto taxpayers. The latter could also come about by design, with pooling arrangements underwritten by government.

Otherwise, premiums must be high enough to absorb the foreseen, but unquantifiable, risks. Then either the insurer makes excess profits, from covering rather than bearing risks, or these profits are returned to policyholders through some form of participation. In fact, the venerable with-profits concept exactly fits the bill. Actuaries in countries in which participation is mandatory might well be wondering what the problem is.

An obstacle to the participating approach is unrestricted freedom to compete on the basis of price, right up to the point at which all future bonus is discounted and we are back to nonparticipating contracts. In theory, such freedom is limited by valuation regulations, so that capital requirements should ensure that adequate premiums are charged. Alternatively, pricing regulations might be preferred, perhaps only for the first slice of cover. This runs against the grain of tradition in the U.K. and in the U.S. but any suggestion deserves to be examined on its merits, in the genetics debate. Paradoxically, just as the Third Life Directive in Europe is removing pricing regulations, minimum tariffs could play an unexpected, but useful, role.

I admit that the problem of adverse selection is not solved just by showing that its costs might be containable, as I have tried to do in the case of life insurance. Many people will, quite reasonably, find it hard to accept that adverse selectors should be allowed their cross-subsidies, even if they are limited. But this can only take place because of the freedom that individuals have to define their own need for insurance

cover. An extreme answer would be to make the first slice of insurance compulsory (rather as third-party motor insurance is compulsory for car owners in the U.K.). Between that extreme and the current extreme lie many possibilities; for example, denying access to various forms of insurance unless the first slices of all the “essential” insurances (life, health, long-term care) are in place (rather as cars cannot be licensed in the U.K. unless proof of insurance is supplied).

Finally, I suggest that simplicity is the key to any solution aimed at the mass market. More sophisticated ideas than those described here might possibly work behind the scenes, as arrangements between insurers and reinsurers, but they are no basis for most peoples’ management of their affairs.

## 9. RESEARCH ISSUES

### 9.1 Public Policy

Insurance companies have always carried out statistical research in the course of their business. However, it has typically been limited to immediate business needs. While this reflects commendable frugality, it perhaps fails to meet wider needs that the debate on genetics highlights:

- a. Insurance research (other than market research) is usually based on data collected in the normal course of business; for example, mortality studies are based on in-force data and claims data. Phenomena whose study would require the collection of other data often go unstudied, even such “classics” as selective lapsing. In the context of discrimination and adverse selection, studies of insurance-buying behavior could not easily be based on available business statistics.
- b. The insurance industry will seem complacent to policy-makers if it regards research as a luxury, and not a necessity. Research does not give perfect or immutable answers; insurance-buying behavior can change, for example. But to advance a policy greatly at odds with much of public opinion, such as the “right to underwrite,” without credible research to back it up (or worse, without even appearing to recognize the need for such research) is to invite a rebuff.

I suggest that the insurance industry needs to take a broad view of the research base needed to underpin any particular practice that affects social, as well as business, policy. To do so, it will have to look outside its own domain.

## 9.2 Insurance Modeling: Research Needs

In Section 2.4h, I listed as a requirement of a useful model that it should be specified in terms of quantities capable, in principle, of being estimated through suitable statistical studies. The remarkable success, and long history, of actuarial mortality studies stems largely from the fact that they rely on easily available business statistics. But, as suggested above, that should not necessarily be a criterion for accepting or rejecting any particular model; it depends on the wider importance of the problem being studied.

The novel feature of models for genetic testing is the introduction of uninsured lives. Conventional actuarial models are conditional, in that it is assumed that insurance has been purchased, underwriting has been carried out, and the risk factors are known. Significant uncertainty about risk factors forces us to use an unconditional model.

The Markov model of Section 3 is specified in terms of transition intensities. These have the statistical merit of being estimated simply as occurrence/exposure rates, if such data can be found (Macdonald 1996a). Alternatively, the integrated intensities can be estimated from cohort data (Macdonald 1996b). I consider the various types of transition in the model, namely, those governing mortality, genetic testing, and insurance-buying behavior.

- a. The main problem in estimating transition intensities representing mortality was discussed in Section 7, namely obtaining sufficient quantities of data classified by genetic type and other relevant factors. If, as is usual, insurance companies do not collect data that they will not (or cannot) use, only external studies will supply these data.
- b. Genetic testing is developing so fast that we cannot say what pattern it might eventually follow. It is reasonable to suppose that medical research will furnish estimates of the population frequencies of genetic disorders (as is already the case with well-known disorders) that will give some guidance on test outcomes. The incidence of testing itself will depend on attitudes and behavior as much as anything else; perhaps some information could be gleaned from carefully constructed panel surveys.
- c. Insurance-buying behavior also depends on attitudes and behavior. Although much market research is carried out, little of it is published, and perhaps even less of it would be relevant. I suggest that the insurance industry might consider what research could be done, in the public interest as well as its own. A useful first step would be to carry out panel surveys of industry practitioners, medical

professionals, and the public, to find out what effect genetic information would have on their desire for insurance, given their respective levels of knowledge.

## 9.3 Other Types of Insurance

Research is even more crucial in health and long-term-care insurance. In both cases, the Markov (or semi-Markov) framework is well established (Waters 1984; Macdonald 1996a; Jones 1997a, 1997b). One area on which researchers could usefully concentrate is Alzheimer disease; it (with other forms of dementia) is one of the major contributors to long-term-care costs, and there are now some well-understood genetic markers.

## 10. CONCLUSIONS

I suggest that the most important contribution actuaries can make to the debate on genetics and insurance is to develop models that delineate clearly the effects of various policy options, and pinpoint where research is needed. Only then can policy-makers make informed choices that give the insurance industry a feasible role.

Of the two modeling approaches most familiar to actuaries, Markov models appear to be more useful than random lifetime models, not least because Norberg's equations (Norberg 1995) make them attractive from a computational point of view. I have developed the Markov model for life insurance used in Macdonald (1997) and have shown that if adverse selection in the form of higher sums assured can be avoided, (1) 10% is a reasonable order of magnitude for the additional costs of adverse selection, and (2) there is little need for stronger risk-based capital margins in respect of the OR class.

In the case of multifactorial disorders, for which the extra mortality might be quite small and subject to other influences, it is not obvious that sufficient data will be available to quantify these differences reliably. It is unlikely that epidemiological studies carried out in the course of genetic research will supply enough data, so the possibility of using many such disorders as rating factors might be limited.

Product design faces the basic problem that the risks posed by genetics are not quantifiable in advance. The insurance industry has a tried and tested strategy to cope with this—participating policies—but the application of that principle to markets that have traditionally been low-cost and nonparticipating might require some limits to be placed on traditional

freedom over tariffs. This might be quite appropriate for the first layer of insurance cover; it will be for policy-makers to decide where the balance lies between the social role and market freedoms of the insurance industry.

I urge the insurance industry to take a broad view of research, and in particular to look beyond its traditional (though cheap) databank of its own business statistics. More outward-looking research is needed to make the boundary between commercial interest and scientific principle clear for all to see.

### REFERENCES

- BECKMAN, P. 1971. *A History of  $\pi$* . Boulder, Colorado: Golem Press.
- BOWERS, N. L., GERBER, H. U., HICKMAN, J. C., JONES, D. A., AND NESBITT, C. J. 1986. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries.
- CHUFFART, A. 1996. *Genetics and Life Insurance in Europe*. Zurich: Swiss Re.
- COLLETT, D. 1994. *Modelling Survival Data in Medical Research*. London: Chapman & Hall.
- CONTE, S. D., AND DE BORG, C. 1972. *Elementary Numerical Analysis*. New York: McGraw-Hill.
- GERBER, H. U. 1990. *Life Insurance Mathematics*. Berlin: Springer-Verlag.
- HARPER, P. S. 1997. "Genetic Testing, Life Insurance, and Adverse Selection," *Philosophical Transactions of the Royal Society B* 352:1063–66.
- HESSELAGER, O., AND NORBERG, R. 1996. "On Probability Distributions of Present Values in Life Insurance," *Insurance: Mathematics & Economics*, 18:35–42.
- HOEM, J. M. 1969. "Markov Chain Models in Life Insurance," *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* 9:91–107.
- HOEM, J. M. 1988. "The Versatility of the Markov Chain as a Tool in the Mathematics of Life Insurance," *Transactions of the 23rd International Congress of Actuaries, Helsinki* S:171–202.
- HOUGAARD, P. 1984. "Life Table Methods for Heterogenous Populations," *Biometrika*, 71:75–83.
- JONES, B. L. 1997a. "Stochastic Models for Continuing Care Retirement Communities," *North American Actuarial Journal* 1, no. 1(January):50–73.
- JONES, B. L. 1997b. "Methods for the Analysis of CCRC Data," *North American Actuarial Journal* 1, no. 2(April):40–54.
- LEIGH, T. S. 1990. "Underwriting: A Dying Art?" (with discussion), *Journal of the Institute of Actuaries* 117:443–531.
- MACDONALD, A. S. 1996a. "An Actuarial Survey of Statistical Models for Decrement and Transition Data I: Multiple State, Poisson and Binomial Models," *British Actuarial Journal* 2:129–55.
- MACDONALD, A. S. 1996b. "An Actuarial Survey of Statistical Models for Decrement and Transition Data II: Competing Risks, Non-Parametric and Regression Models," *British Actuarial Journal* 2:429–48.
- MACDONALD, A. S. 1997. "How Will Improved Forecasts of Individual Lifetimes Affect Underwriting?" *British Actuarial Journal* 3, no. 15(December):1009–25.
- MOULTRIE, T. A. AND THOMAS, R. G. 1997. "The Right to Underwrite? An Actuarial Perspective with a Difference" (with discussion), *Journal of Actuarial Practice* 5, no. 1:125–47.
- NORBERG, R. 1995. "Differential Equations for Moments of Present Values in Life Insurance," *Insurance: Mathematics & Economics* 17:171–80.
- O'NEILL, O. 1997. "Genetic Information and Insurance: Some Ethical Issues," *Philosophical Transactions of the Royal Society B*, 352:1087–93.
- POKORSKI, R. J. 1996. "Use of Genetic Tests to Predict and Diagnose Cancer: An Insurance Perspective," *Journal of Tumor Marker Oncology* 11:33–44.
- POKORSKI, R. J. 1997. "Medical Underwriting in the Genetics Era," *Journal of Insurance Medicine* 29:107–19.
- PRITCHARD, D. J. 1997. *Life Assurance: Financial Implications of a Change in Insuring Behaviour Resulting from Individuals' Increased Knowledge of Their Genetic Predispositions*. M.Sc. dissertation, Heriot-Watt University, Edinburgh.
- TAN, K. W. 1997. *The Financial Impact of Genetic Testing on Annuities*. M.Sc. dissertation, Heriot-Watt University, Edinburgh.
- VAUPEL, J. W. 1988. "Inherited Frailty and Longevity," *Demography* 25:277–87.
- VAUPEL, J. W., MANTON, K. G., AND STALLARD, E. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography* 16:439–54.
- WATERS, H. R. 1984. "An Approach to the Study of Multiple State Models," *Journal of the Institute of Actuaries* 111: 363–74.

### APPENDIX

#### NORBERG'S EQUATIONS FOR MOMENTS OF POLICY VALUES

The calculation of moments in the Markov model follows Norberg (1995). We adapt Norberg's notation slightly. Suppose that a continuous annuity is payable at rate  $b_{x+t}^{ij}$  during sojourns in state  $ij$ , and that on transition from state  $ij$  to state  $ik$  at age  $x + t$ , a sum assured of  $b_{x+t}^{ik}$  is payable. Let  $V_{x+t}^{(q)ij}$  be the  $q$ -th moment about zero of the discounted future cash flows at time  $t$ , given that a life in the  $i$ -th subgroup is in state  $ij$  at age  $x + t$ . These moments satisfy the following system of differential equations:

$$\frac{d}{dt} V_{x+t}^{(q)ij} = (q\delta + \mu_{x+t}^{ij})V_{x+t}^{(q)ij} - qb_{x+t}^{ij}V_{x+t}^{(q-1)ij} - \sum_{k \neq j} \mu_{x+t}^{ijk} \sum_{r=0}^q \binom{q}{r} (b_{x+t}^{ijk})^r V_{x+t}^{(q-r)ik}, \tag{7}$$

where  $\mu_{x+t}^{ij} = \sum_{k \neq j} \mu_{x+t}^{ijk}$ . The case  $q = 1$  furnishes the familiar Theile's equations. The system can be solved recursively, using as boundary conditions that all terminal policy values are zero. We used a fourth-order Runge-Kutta procedure (Conte and de Boor 1972).

Define  $V_{x+t}^{(q)ij}$  to be the  $q$ -th moment about zero of the discounted future cash flows, given that at time  $t$  the life is known to be in one of the states  $1j, 2j, \dots, Mj$ , but that  $i$  is unknown. In other words, we know whether or not a life is insured and whether or not a life has taken a genetic test, but we do not know the subgroup to which the life belongs. Then,

$$V_{x+t}^{(q)j} = \sum_{i=1}^M p_i V_{x+t}^{(q)ij}. \tag{8}$$

As in Norberg (1995) we then calculate the  $q$ -th central moments, denoted  $m_{x+t}^{(q)j}$ , from

$$m_{x+t}^{(q)j} = \sum_{p=0}^q \binom{q}{p} (-1)^{q-p} V_{x+t}^{(p)j} (V_{x+t}^{(1)j})^{q-p}. \tag{9}$$

## DISCUSSION

**JAMES C. HICKMAN\***

### INTRODUCTION

That we live in an information age is a cliché. Yet in common with many clichés, this one casts a shadow of truth. The concomitant growth of basic science, statistical agencies—both public and private—that collect and summarize data, and the technical means to store and transmit information has changed our society, our economy, and our individual lives.

Insurance classification is an information-processing operation, and it is being profoundly changed in the information age. Brockett et al. and Macdonald have mapped for us one particular area of insurance classification in which new information is clashing with long-held traditions and business processes. The area, defined by the intersection of the domains of

four topics, is a fascinating area for actuaries to explore, an exploration that will yield knowledge useful far beyond the original area of inquiry. Most of the issues discussed, and the models proposed, will be relevant to exploring the implications of the use of new insurance classification information of any type.

The four defining ideas or developments are as follows:

a. *Developments in the Study of Human Genetics.* The first half of this century has been called the age of chemistry. Developments in chemistry influenced the growth and decline of industries and shaped the goals of diplomacy and how wars were fought. The study of chemistry was propelled by the development of the periodic table. This rectangular display of information about the elements is well-known to all chemistry students. An element's position in the table discloses much about its properties and its relationships with other elements. The results of the Human Genome Project can be compared with the periodic table. The display will be many orders of magnitude larger, but its existence will organize knowledge and establish an agenda for research on genetic configurations and human attributes. The impact on research in medicine, sociology, and psychology will likely be comparable to that of the periodic table, and the atomic structure on which it was based, on chemistry and physics.

b. *The Triumph of Open Markets.* As Adam Smith would have predicted, the internal contradictions of communism led to its collapse. The efficiency and self-correcting features of open market economies are now almost universally recognized. Open markets are just that: open. They depend on the free flow of economic information. We should not expect the benefits of efficiency from open markets unless the prerequisite of a rough symmetry of information among the market participants is met. In particular, this is true of insurance markets.

c. *The Right of Privacy.* Privacy is not a civil right enumerated in the first ten amendments to the U.S. Constitution. Yet many judges see a component of privacy in other individual and property rights. The conflict between the implied right of privacy and the open market requirement for symmetry of information is probably the most enduring fault line in insurance law. Marine insurance developed rapidly in England in the eighteenth century. Horwitz (1992) states that the fundamental legal result of that development was "any failure to disclose a material fact that affected risk was sufficient to void an insurance policy." Therefore, in the early days of insurance the requirement

\*James C. Hickman, F.S.A., A.C.A.S., Ph.D., is Emeritus Professor and Dean, University of Wisconsin, School of Business, 975 University Avenue, Madison, Wisconsin 53706.

for symmetry of information was enforced, and at least for shipowners the right of privacy was restricted. It is important to acknowledge that a genetic propensity to Alzheimer's disease is not equivalent to a leaky bilge pump.

d. *The Concept of "Fairness."* The word "fair" is one of the few in the dictionary that has listings for all the principal parts of speech. It is a difficult concept to define, and the meaning has not been invariant across time. For example, the amount of physical contact that is "fair" for NBA players has changed significantly. Yet the concept of "fairness" is powerful. One cannot expect easy success in attempts to change a public or business policy against the current consensus about what is "fair."

## INSURANCE

Brockett et al. were wise to start their paper with a discussion of the definition of insurance. Many discussions of insurance classification, for example, those about health insurance for victims of the HIV epidemic, such as Daniels (1990), are inconclusive because the participants start with different, but unstated, definitions of insurance. It would be presumptive to attempt to end the long discussion of this definition, but Brockett et al.'s sentence in which they quote Dieke seems consistent with the motivation for the models in Macdonald's paper: "The actuary views insurance in terms of the probabilistic law of large numbers and the central limit theorem as 'a mechanism for transferring actuarial risk and dealing with it through pooling.'"

This definition has the advantage of being rooted in the mathematical concept of randomness. It also helps make the limits of insurance more precise. Insurance is not the solution to the inconvenience of financial losses from all circumstances. There may be very powerful reasons for private and voluntary, or government and compulsory, programs to facilitate savings or for government or charitable programs to transfer wealth that are not technically insurance programs. It is likely that there may be losses associated with genetic related diseases that are not amenable to insurance solutions.

## COST OF INFORMATION

Even in an information age, information comes with a cost. There is a practical business-imposed bound on how much risk classification information, genetic or traditional, that the insurer can afford to buy.

A simple model to illustrate this idea is taken from my discussion of Promislow (1987). I examine this issue using elements of statistical decision theory. The development is based on a simple model employing the following notation:

$X$  denotes the random variable that can be interpreted as the present value of future claims on a policy.

$\theta$  denotes a vector parameters that determine the distribution of  $X$ . In life insurance the elements of  $\theta$  might be the numerical values representing age, sex, height, weight, and so on.

$I_j$  denotes an information-gathering process for learning about  $\theta$  that results in the insurance applicant's being assigned to a rating class. We could think of  $I_j$  as a member of a set of classification processes. The process  $I_j$  will not necessarily provide perfect information about  $\theta$ . The information-gathering process assigns a rating class to each applicant.

Before the classification information is collected by process  $I_j$ , the risk parameters have a distribution that can be interpreted as the prior (before the collection of information) distribution of risk parameters in the population of insurance applicants. After the information specified by process  $I$  is collected, the risk parameters have a posterior distribution appropriate for the applicant.

These ideas lead to the following results:

The benefit premium is

$$H(I_j) = E_{X|I_j}[X|I_j] = E_{\theta|I_j} E_{X|\theta}[X|\theta] = E_{\theta|I_j}[H(\theta)]$$

and a typical measure of risk is

$$\text{Var}(X|I_j) = \text{Var}_{\theta|I_j}(E[X|\theta]) + E_{\theta|I_j}[\text{Var}(X|\theta)].$$

If  $I_j$  yields perfect information about  $\theta$ ,

$$H(I) = H(\theta),$$

$$\text{Var}_{\theta|I_j}(E[X|\theta]) = 0,$$

$$E_{\theta|I_j}[\text{Var}(X|\theta)] = \text{Var}(X|\theta),$$

and

$$\text{Var}(X|I_j) = \text{Var}(X|\theta).$$

The insurer would seek to minimize  $\text{Var}(X|I_j)$  over the set of possible information-gathering processes, subject to the budget restriction that the cost of the information-collecting process  $C(I_j) \leq K$ , where  $K$  is the maximum permitted by business conditions for classification expenditures.

Since the risk premium to compensate the insurer for accepting risk is likely to be a nondecreasing function of  $\text{Var}(X|I_i)$ , the process of minimizing  $\text{Var}(X|I_i)$ , subject to the budget restriction, also will reduce the risk premium component of contract premiums. Information, including genetic information, may change premiums charged individuals, but it will reduce the general level of contract premiums by reducing the risk component.

### CLASSIFICATION AND THE LINEAR STATISTICAL MODEL

Linear statistical models have been used to evaluate the efficiency of alternative classification systems. The extension of this application to genetic information is immediate, but to those who have not seen the basic application, it is worth repeating.

For each risk classification method, we estimated by least squares a model of the form

$$Y_i = X_i\beta + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $i$  indexes the  $n$  individuals that belong to the sample of data used in estimating the model,  $\beta$  is a  $p \times 1$  matrix of parameters (risk weights) to be estimated,  $X$  is an  $n \times p$  matrix of the independent variables included in the model (the risk group dummy variables),  $Y_i$  are the actual present value of claims for an individual, and  $\varepsilon$  is a  $1 \times n$  vector of disturbances/random error terms.

The composition of the  $X$ , or design, matrix depends on the risk factors that enter a particular classification system. For example, a set of dummy variables (0 or 1) might be used to code an age and sex classification system. Other dummy variables (0 or 1) might be related to the presence or absence of a genetic mutation. Interaction terms can also be introduced. Each pair of a dataset and a classification system yield a coefficient of determination, or  $R^2 = (\text{sum of squares due to regression})/(\text{sum of squares from overall average})$ . In underwriting this  $R^2$  is also called the *efficiency index* of the classification system. If the standard statistical assumptions about the vector of random error terms seem realized, standard tests of hypotheses about the  $\beta$  parameters (risk weights) can be done.

The model is discussed by Woll (1979) and has been used in health insurance not only to compare classification systems but also to design capitated compensation systems for health care providers to pay higher compensation to providers with patients with higher expected costs.

The results of a major study of health risk adjustment methods, using the linear model in the evaluation process, are reported by Dunn et al. (1995). Their purpose was to evaluate a set of alternative health risk adjustment methods for creating a system of compensating payments among health insurers to minimize the effects of competition on classification standards. In this application, some of the elements of the  $\beta$  vector are related to demographic variables and others with a mapping of ICD9 diagnostic codes, reported for a particular insured individual in the past year, into a set of much lower dimension of summary classes.

Rather surprisingly, in most applications the linear model is used to evaluate competing classification system rather than to construct a new system. A team approach with an experienced applied statistician to test residuals and suggest interaction terms working with a subject area specialist would seem to offer the best chance of success. This would seem especially true with genetic information, where it is well-known that polygenic disorders—interaction terms—are common.

### THE RANDOM LIFETIME MODEL

In Section 4 of his paper Macdonald introduces the random lifetime model and provides reasons why the Markov model is both mathematically convenient and more realistic. The realism argument is particularly powerful for disability and long-term-care insurance in which repeated transitions among states is possible.

The ideas in Macdonald's Section 4 are closely related to those introduced to North American actuaries by Jones (1998). Both Jones and Macdonald introduce the concept of human frailty, which is assumed fixed at birth. In the past frailty could not be observed directly. The progress in genetic testing may permit more direct observation of frailty. Frailty provides a framework for introducing heterogeneity into mortality models, and it was the estimation consequences of this heterogeneity that Jones studied.

Jones employs a survival function that depends on a frailty value; that is,

$$s(t, z) = \Pr(T > t | Z = z)$$

and

$$\mu(t, z) = \frac{-\partial s(t, z)}{\partial t} / s(t, z).$$

Jones's example is  $\mu(t, z) = z \mu(t)$ . This can be compared to Formulas (5) and (6) in Macdonald's paper. Both Jones and Macdonald assume the frailty, denoted

by  $Z$ , is a random variable with p.d.f.  $f_z(z)$ . The examples differ in that Jones assumes that  $Z$  is a continuous random variable, while Macdonald assumes a discrete distribution for  $Z$ .

To provide a framework for examining the effect of genetic testing, we modify Jones's notation. A subscript  $I_j(X)$  is added to the survival function and the force of mortality, and we now have the survival function

$$s_{ij(x)}(t, z)$$

and

$$\mu_{ij(x)}(t, z),$$

where  $I_j(X)$  denotes the risk class to which the insured is assigned on the basis of information process  $j$ , and observing vector  $X$  of information. Without classification,  $I_j(X)$  might be age  $X$ .

Law, tradition, regulation, and business practice may limit the variables that can belong to the information vector. The effect of genetic testing information, if any, is assumed captured in the distribution of  $Z$  given the value  $I_j(X)$ .

The  $q$ -th moment of the loss variable associated with a whole life policy, with continuously paid premium rate  $\pi(I_j(X))$ , is determined from

$$\int_0^\infty \int_0^\infty [e^{-\delta t} - \pi(I_j(X))\bar{a}_t]^q f_{T|Z, I_j(X)}(t) f_{Z|I_j(X)}(z) dz dt.$$

If genetic information is available, the distribution of frailty  $Z$ , given information  $I_j(X)$ , may be concentrated at a single point. It is natural to assume that the rate  $\pi(I(X))$  is determined by the equivalence principle, that is, by equating the first moment of expected losses to zero. If there are restrictions on the information that can be used in premium determination, the variance of the loss variable might increase because of the resulting bias term. Bias would result if  $\pi(I_i(x))$  is used as the premium because of regulation or error, when the distribution of  $T$  using more relevant information would be based on  $I_j(X)$ ,  $i \neq j$ .

Within this model the effect of antiselection would appear in the distribution of  $Z$ . Probability density would shift from low values of  $z$  to high values of  $z$ . The effect of this shift might be countered by the insurer by selecting a different information and classification function,  $I(x)$ , to retain the desired distribution of  $Z|I_j(X)$ . This game, with applicants shifting the

distribution of  $Z|I_j(X)$ , and the insurer adjusting  $I_j(X)$  to some  $I_j(X)$ ,  $i \neq j$ , might continue until a new premium, determined by the equivalence principle, would result in an economic equilibrium influenced by the cost of information.

In these calculations approximate integration, possibly using Monte Carlo methods, would be the principal tool. The measurement of expected losses, biases in statistical language, or changes in risk due to antiselection will require making assumptions about the distribution of  $Z|I(x)$  forced by insurance applicants in seeking their self-interest. These shifts might be estimated using research as outlined by Macdonald in Section 9.2 or illustrated with subjective-based assumptions.

### Example

A simple example, building on one of Jones's, may help clarify these ideas. We define two p.d.f.s.

$$f_{T|Z, I_j(X)}(t) = e^{-\lambda z t}, t, \lambda, z > 0,$$

$$f_{Z|I_j(X)}(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z}, z, \alpha > 0.$$

The properties of the gamma distribution for  $Z$  are listed by Jones,  $E[Z|I_j(X)] = 1/\alpha$  and  $\text{Var}(Z|I_j(X)) = 1/\alpha$  and as  $\alpha \rightarrow \infty$ , the distribution approaches a single-point distribution at  $z = 1$ . The predictive distribution, used in moment calculations of insurance loss variables, is

$$\begin{aligned} f_{T|I_j(X)}(t) &= \int_0^\infty f_{T|Z, I_j(X)}(t) f_{Z|I_j(X)}(Z) dz \\ &= \frac{\lambda \alpha^{\alpha+1}}{(\alpha + \lambda t)^{\alpha+1}}. \end{aligned}$$

The predictive distribution combines the variability of random lifetime and frailty. The predictive survival function of  $T$ , denoted by  $\bar{s}(t)$  by Jones, is in our example

$$\bar{s}_{I_j(X)}(t) = \frac{\alpha^\alpha}{(\alpha + \lambda t)} \alpha.$$

This can be compared with the conditional survival function of  $T$

$$s_{Z, I_j(X)}(t) = e^{-\lambda z t}.$$

The predictive variance of  $T$  is given by

$$\begin{aligned} \text{Var}(T) &= \text{Var}(E[T|Z]) + E[\text{Var}(T|Z)] \\ &= \text{Var}\left(\frac{1}{\lambda Z}\right) + E\left[\frac{1}{(\lambda Z)^2}\right] \\ &= \left(\frac{1}{\lambda^2}\right) \{\text{Var}(Z^{-1}) + E[Z^{-2}]\} \\ &= \left(\frac{1}{\lambda^2}\right) \{2E[Z^{-2}] - E[Z^{-1}]^2\} \\ &= \left(\frac{1}{\lambda^2}\right) \left[\left(1 - \frac{1}{\alpha}\right)^{-2} \left(1 - \frac{2}{\alpha}\right)^{-1}\right], \quad \alpha > 2. \end{aligned}$$

As  $\alpha \rightarrow \infty$ , which in this example is equivalent to increasing the precision with which frailty is known, the predictive variance approaches  $\lambda^{-2}$ .

In applications, the distribution of  $Z|I_j(X)$  is influenced by the distribution of frailty in the population and the self-selection of applicants for insurance. The self-interest of rational insurance applicants would lead to a prediction that the probability of an applicant with a high frailty value will be higher than the occurrence in nature of this value.

**PARTICIPATING SOLUTION**

In Section 8 Macdonald presents persuasive arguments for using participating policies as the natural response to the challenge of new genetic testing information. An additional argument supporting his recommendation is the likelihood that once a correspondence is established between a genetic configuration and human mortality or morbidity, a cell-specific therapy will be developed. Genetic information will not increase mortality, and it is very likely to reduce it at an uncertain future time. A natural market response to this possibility is a participating policy with the ability to adjust the mortality component of the policy dividend as information becomes available.

**SUMMARY**

1. Brockett et al. and Macdonald have provided an accurate map to guide actuaries through the new territory of genetic testing.

2. We should not expect the benefits of an efficient insurance market without symmetric information. In some situations nonmarket solutions may be required to meet a societal need.
3. Information has costs. Markets will balance the value of genetic information with its cost. Because of cost or difficulty in estimating its effect on the distribution of insurance claims, much genetic information will be economically irrelevant.
4. Risk premiums within insurance premiums will decline with genetic information.
5. There are at least two models for studying the effect of antiselection based on genetic information on life insurance: the Markov and random lifetime models.
6. Participating policies exist in the market and can provide the flexibility to manage the uncertainty in genetic information of most types.

**REFERENCES**

DANIELS, N. 1990. "Insurability and the HIV: Epidemic Ethical Issues in Underwriting." *Milbank Quarterly*, 68:497-525.

DUNN, D. L., ROSENBLATT, A., TAIRA, D. A., LATIMER, E., BERKTO, J., STOIBER, T., BRAUN, P., AND BUSCH, S. 1995. *A Comparative Analysis of Methods of Health Risk Assessment, Final Report*. Schaumburg, Ill.: Society of Actuaries.

HORWITZ, M. J. 1992. *The Transformation of American Law, 1780-1860*. New York: Oxford University Press.

HICKMAN, J. C. 1987. Discussion of S. D. Promislow, "Measurement of Equity." *Transactions of the Society of Actuaries*, 39:234-46.

JONES, B. L. 1998. "A Model for Analyzing the Impact of Selective Lapsation of Mortality." *North American Actuarial Journal*, 2:79-86.

WOLL, R. G. 1979. "A Study of Risk Assessment." *Proceedings of the Casualty Actuarial Society*, 66:84-138.

*Additional discussions on this paper can be submitted until July 1, 1999. The author reserves the right to reply to any discussion. Please see the Submission Guidelines for Authors for instructions on the submission of discussions.*