GENETICS AND INSURANCE: WHAT HAVE WE LEARNED SO FAR?

By Angus Macdonald

ABSTRACT

Genetics and insurance is an area unusually exposed to rapid scientific advance, close public and political scrutiny, and popular myth. It may be leading the way towards evidence-based underwriting. This survey paper describes some of the experience gained since actuarial involvement began in the mid-1990s, particularly the vital link with genetic epidemiology. We survey the relevant aims and outputs of genetic epidemiology, mainly relating to single-gene disorders, and the use of genetic epidemiology in actuarial models. The part that actuarial models might play in evidence-based approaches to underwriting and policy-making is discussed.

KEYWORDS

Adverse Selection; Discrimination; Genetic Epidemiology; Genetics; Insurance; Underwriting

CONTACT ADDRESS

A. S. Macdonald, B.Sc., Ph.D., F.F.A., Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3209; Fax: +44(0)131-451-3249; E-mail: A.S.Macdonald@ma.hw.ac.uk

1. Introduction

Human genetics has developed at an astonishing pace since 1990: the sequencing of the human genome has gone from an idea to near fruition, most of the major disease-causing genes have been located, and attention is now turning to proteomics, deducing protein function from gene sequence. But amid all the justified excitement, it is easy to overlook the fact that this amazing progress has been made in certain specific areas of genetical work; mostly, those that can be pursued in laboratories, or inside computers, where lavish funding can lead to quick results. Genetics is a larger subject, and the faster molecular genetics races ahead, the more work it leaves behind it, yet to be done.

This is crucial to understanding where genetics stands in relation to insurance, because insurance issues revolve around a discipline — genetic epidemiology — that is among those following in the wake of the laboratory science. Molecular genetics may be about to tell us why, for example, certain mutations in the Presenilin-1 gene lead to overproduction of long-chain β-amyloid, and why slightly longer versions of these molecules are more likely to agglomerate in peoples' brains, hence 'explaining' early-onset Alzheimer's disease (EOAD), but it does not tell us how likely it is that EOAD will appear before age 30, or 40, or 50. The literature of molecular genetics has hundreds of references to Presenilin-1; the literature of epidemiology has almost none. A recent survey (Dartigues & Letenneur, 2000) concluded with a mere one paragraph on all three genes known to lead to EOAD. Yet these genes were included in the list of genes regarded as significant for insurance by the Association of British Insurers (the ABI, ABI (1999)).

If the miraculously precise science of genetics lies at one end of a scale, the very human topic of ethics lies at the other, part of the so-called Ethical, Legal and Social Implications (ELSI) program. This deals with the outcomes of genetics for people and families, in terms of self-knowledge and living with the consequences, and worries about insurability figure prominently. Insurance is perhaps the most quantitative aspect of the ELSI program, so the lack of a strong quantitative input to the discussion appears surprising, until we remember that the necessary knowledge must come from genetic epidemiology.

Section 2 introduces some basic facts about human genetics, mainly to help to make sense of genetic epidemiology. The main aims of this paper are to describe, to actuaries experienced in actuarial mortality analyses, the different problems facing genetic epidemiologists (partly in the hope that actuaries might make a useful contribution); we do this in Section 3, and then Section 4 surveys applications of genetic epidemiology in actuarial modelling. Section 5 discusses briefly the rôle of actuarial models in the wider policy-making arena, and what that means for actuaries in both commerce and academia.

2. Human Genetics

This section introduces the bare minimum of human genetics needed to appreciate insurance related issues. A good introduction for insurance professionals is Fischer & Berberich (1999), and standard texts are Pasternak (1999) and Strachan & Read (1999).

When all goes well, genes have three functions:

(a) controlling the production and/or processing of *all* the substances involved in *every* biochemical process in the body;
(b) passing on exact copies of themselves during cell division, by which means a fertilised egg develops into an adult, and tissues are repaired during life; and
(c) passing on exact copies of half of their number to offspring, where they will be joined by half of the other parent's genes.

A gene can be regarded as a region of DNA that encodes for the production of a protein or other molecule involved in one or more processes or *pathways* in the body. It can be switched on and off, to a different extent in the cells of different organs, as more or less of the substance is needed by those organs; this switching or *regulation* is itself controlled by substances produced by other genes that are switched on and off as required, by substances produced by other genes . . . and so on. The complexity of these regulatory networks can be immense.

(c) above leads to Mendel's rules for the inheritance of single-gene disorders. Everyone has two copies of (almost) all genes — we will ignore the exceptions — one inherited from each parent. Each performs its rôle in the cell when required, so the 'normal' amount of gene product is that produced by two genes. Therefore, what happens when (a) above goes wrong, because one or both parents has passed on a faulty (mutated) version of a gene?

(a) If the mutation produces a toxic substance, inheriting one copy may be sufficient to cause disease or death. Selection pressure tends to eliminate such mutations from the gene pool, so they are rare, and it is reasonable to ignore the possibility that both parents will carry such severe mutations in the same gene; therefore if one parent is

a carrier, each of their offspring has a 50% chance of inheriting it. This is Mendel's rule for *dominantly inherited traits*.

(b) If the mutation produces a more or less harmless but useless substance, and one 'normal' copy of the gene can produce enough of the correct substance, then inheriting just one mutation might not lead to disease (although it might lead to some deficiency). But inheriting two mutations could eliminate a vital function, and a *recessively inherited trait* will result. Mendel's rules then are:
   (1) if one parent is an unaffected carrier (has one mutation, often unaware of the fact), 50% of the offspring will be unaffected carriers, and 50% non-carriers; and
   (2) if both parents are unaffected carriers, 25% of their offspring will be affected, 50% will be unaffected carriers and 25% will be non-carriers.

Mendel's rules govern the assessment of the risk of an as-yet unaffected person in a family affected by a dominant mutation being a mutation carrier, hence, implicitly, underwriting based on a family history of a dominant Mendelian disorder.

However, rather few gene disorders seen in adults are so drastic as to lead definitely to severe disease, still fewer with onset deferred to older (economically active) ages; the same evolutionary pressure that selects advantageous mutations for survival and reproduction tends to eliminate disadvantageous mutations, so that those that are observed in modern adults tend to be:

(a) mutations that lead to the accumulation of damage or noxious substances, but so slowly that it would barely have affected human evolution until modern times (dominant disorders like Huntington's disease (HD) or EOAD);
(b) mutations in *oncogenes* in which one functioning copy of a gene suffices to prevent onset of cancer, for example by repairing damaged DNA, so that a mutation in that functioning copy in just one cell in adult life robs that cell of its protection (familial breast cancer (BC) and ovarian cancer (OC)); or
(c) mutations that confer an *advantage* if only one copy is inherited (for example, the sickle cell and thalassaemia mutations offer protection against malaria) but do not support normal function if two are inherited, leading to recessive disorders in adults.

These are just the 'simple' genetic disorders, traceable to a single gene (or to several genes but each alone sufficient to cause the disorder) with a Mendelian pattern of inheritance. By far the majority of the genetic contribution to disease, however, is *multifactorial*: slight variants (polymorphisms) of many genes in complex metabolic and regulatory pathways interacting with the environment and lifestyle to play a part in common, complex disorders. These are poorly understood and will not concern us much in what follows.

3. Genetic Epidemiology and Survival Analysis

3.1 *What Questions Do Actuaries Want To Answer?*

Actuaries address questions about insurance risk. Since insurance contracts may, in principle, be sold to people of any age, extending over any term, actuaries need quite detailed age-dependent risk estimates. The simplest such model meeting this need, that is

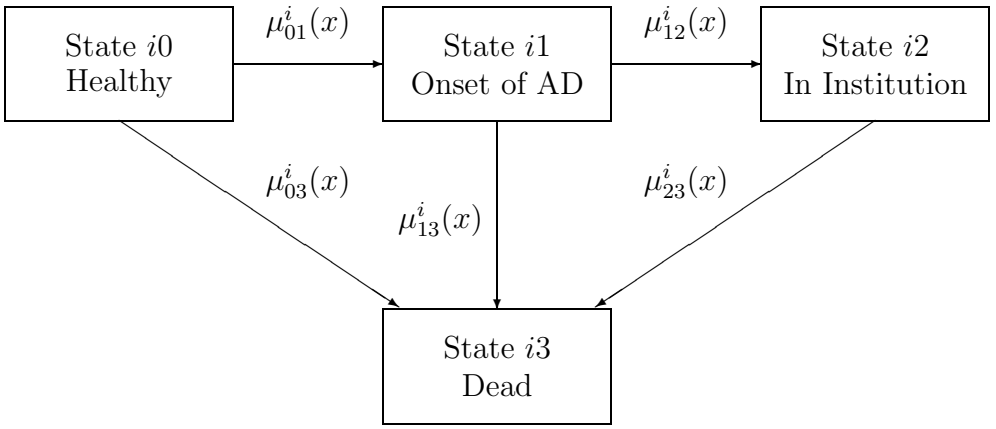

Figure 1: A Markov model of Alzheimer's disease, for a person with the ith of several relevant genotypes.

flexible enough to accommodate both insurance and genetical questions, is the continuous-time multiple-state model. We shall tacitly assume this to be in the background until we discuss it further in Section 4. As an example, Figure 1 shows a model of certain events associated with the progression of Alzheimer's disease (AD). The model is parameterised by its transition intensities, which for simplicity we assume to depend on age alone (a Markov model). The problem, that the actuary hopes that the genetic epidemiologist has already solved, is the estimation of those intensities that depend on genotype; in this case, rates of onset of AD at least.

3.2 *What Questions Do Genetic Epidemiologists Try To Answer?*

It is helpful to start by trying to understand what interests genetic epidemiologists. Much as we might like to analyse raw data ourselves, this is usually impossible. The data will have been collected in the course of medical research and, by its very nature, data for genetical studies often include details of whole families, whose privacy is important. Its dissemination is therefore (rightly) usually limited to researchers whose projects have a clear beneficial purpose, and insurers are not generally regarded as such. Therefore, our understanding of the implications of genetics must often rest on what is published in the medical literature, which addresses a different set of questions. The following list is limited to major questions of common interest to actuaries and geneticists.

(a) *What genes are associated with what diseases?* Gene-hunting means finding the location, on the genome, of a gene associated with a disease. It has been hastened by the evolution of the techniques used to sequence the genome (Strachan & Read, 1999), but it still requires particular variations in the 3 billion or so 'letters' of the genetic code (maybe as small as a single change) to be accurately matched to the occurrence of the disease. The hunts for some of the genes causing major Mendelian disorders were true epics of international science (Davies & White, 1995), and these were the 'easy' ones because family histories made it plain that the genes were there to be

found. Finding mutations of modest or low penetrance (see below) that may be heterogeneous requires statistical approaches to identify the locus of greatest likelihood (Lange, 1997; Sham, 1998).

(b) *What is the penetrance of a mutation?* Penetrance means the probability that a mutation carrier would develop the disease by age x, if all other decrements were absent. We denote it $q(x)$. It is similar to the so-called independent probability of decrement in the single-decrement table associated with one of the decrements in a multiple-decrement table (Neill (1977), although modern statistical critiques of this traditional actuarial approach mean that it ought to be deprecated, see Crowder (2001) for example). If μ_x is the rate of onset (transition intensity) in a multiple-state model intended for actuarial use, we have:

$$q(x) = 1 - \exp\left(-\int_0^x \mu_t dt\right)$$

so penetrance estimates are very useful, if they are available. Even among the severe single-gene disorders, penetrance often falls well short of 100% at the highest ages.

(c) *How common is a mutation?* The burden of disease results from the combination of the frequency of a mutation (in the population), its penetrance, and the course of the disease itself. Estimates of mutation frequency are therefore needed. For insurance application, these are important for estimating the potential for adverse selection. They often have to be estimated at the same time as penetrance estimates, and penetrance of less than 100% makes this a difficult task. The migration long ago of ancestors who 'founded' modern population groups means that mutation frequencies may be exceptionally high among identifiable groups; BRCA1 mutations among Ashkenazi Jewish women is an example (BRCA is an abbreviation of 'breast cancer'). Since these groups may be intensively studied, for the obvious reason that 'interesting' genes are present, it is necessary to be cautious in applying published estimates more widely.

3.3 *The Ideal Case: Survival Analysis*

The ideal study of demographic rates would begin with a very large population of individuals known to be completely homogeneous and independent (in the usual sense) and would follow them and measure precisely the time(s) at which the event(s) of interest took place. Of course this is impossible, and actuaries and statisticians have come to accept many compromises to accommodate what is possible. The resulting subject of survival analysis now extends far beyond its ancient subset which, by and large, still suffices for most actuarial work.

For a genetical study, addressing an apparently simple question, like the effect of mutation X in gene Y on the onset of disease Z, we might accept a population homogeneous in respect of mutation X, and a control population lacking mutation X. We might include a reasonable number of potential covariates, such as sex and smoking habits. We would expect some observations to be censored and would try to recruit enough subjects to allow for this. We could then hope to apply standard methods of survival analysis, though we would not be too surprised if unforeseen heterogeneity or missing data made the analysis harder than planned.

As far as much of genetic epidemiology is concerned, the sort of study described above, warts and all, would appear to be absurdly straightforward. The main problems (but by no means the only problems) are that:

(a) finding a population "homogeneous in respect of mutation X" is usually impossible; and
(b) there is often not a simple causal relation between mutation and disease.

3.4 *Reality: Genetical Data*

The idea of isolating a population of known mutation carriers depends on being able to say who is or is not a mutation carrier. Before DNA-based genetic tests were available, the only way to do so was to wait for some clinical manifestation of the mutation to appear (often but not always the disease of interest). The problem then was how to treat censored observations? For example, let $p(x)$ be the probability that a *known* carrier of a HD mutation will survive to age x free of symptoms. Suppose a healthy woman aged 40 had a father who developed HD. Then the probability that she carries the HD mutation is not 1/2, which is P[Mutation carrier | Alive at age 0] but:

$$\text{P[Mutation carrier | Alive and free of HD at age 40]} = \frac{p(x)}{1+p(x)}. \tag{1}$$

Suppose the study ended when this woman was age 40. We cannot just omit her (and others like her) from the analysis, because that would systematically ignore the mutation carriers most 'resistant' to the onset of HD, a clear bias. It follows that in any family, that we include in a study because some of its members have had HD, we must also include those family members who have *not* had HD, some of whom never will. The basic unit of analysis is not the individual recruited to a study, but their entire family. The sort of information we might have is a collection of pedigrees, see Figure 2 (from Gui & Macdonald (2002)) for example.

Therefore, until DNA-based tests appeared, the very starting point of conventional survival analysis — a group homogeneous for the risk factor of interest — was unattainable. Now that DNA-based tests are available for many genetic disorders, it might be imagined that this problem is solved, but this is not so, and in some respects matters are worse. It is not easy to obtain the samples needed for genetic testing from *all* the members of *all* the families recruited to a study. The more severe and untreatable the disease, the more likely it is that some people will prefer not to be tested, or to take part in the study. The prevalence of testing for HD, for example, is only about 10–20% (Meiser & Dunn, 2000). So for some time to come, epidemiologists might have to study populations that contain both tested and untested individuals, which might be a harder problem than if everyone were untested.

From this springs the second distinctive problem of genetic epidemiology, mentioned in Section 3.3. If we do not have an entirely tested population, then often we will have to infer the presence of a mutation from the occurrence of the disease itself. This can be quite reliable in some cases, like HD, where the disease has *no other cause* but the gene mutation. Otherwise, it is only possible to infer some probability of carrying a mutation, much more complicated than the Mendelian probabilities of inheritance. For example

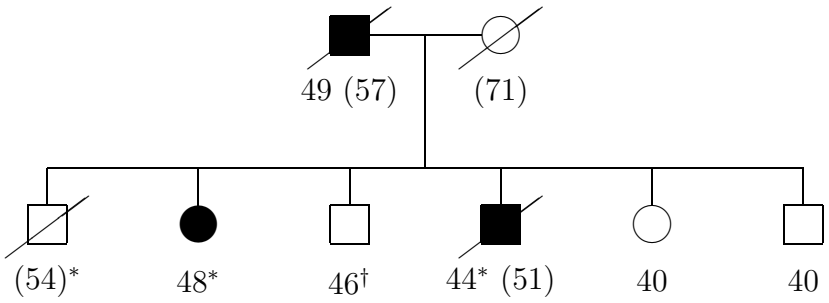

Figure 2: A hypothetical example of a pedigree. Squares are males, circles are females, and a slash denotes death. Affected individuals are shown as filled squares/circles. The age at onset or oldest observed age free of disease is shown, and age at death is given in brackets. An asterisk means that a person has been tested and does carry a mutation, a dagger that he/she has been tested and does not carry a mutation. By convention siblings are listed left-to-right in birth order. Source: Gui & Macdonald (2002).

mutations in the BRCA1 and BRCA2 genes are associated with BC, but only about 5% of BC is associated with genes with a recognisably Mendelian pattern of inheritance. The existence of a family history of BC, even if there are several cases, is not necessarily strong evidence of a single gene at work.

To these basic complications, we can add the following list of difficulties, all more or less formidable:

(a) *One gene, multiple endpoints.* Some genes are associated with several diseases, not just one; for example BRCA1 mutations can lead to breast or ovarian cancer. Therefore a family history of one disease can, *via* a shared genetic cause, imply increased risk of another disease.

(b) *One gene, multiple mutations.* Some genes have just a few, more or less common, alleles; for example there are three common alleles of the Apolipoprotein E (APOE) gene of which one, $\varepsilon 4$, is associated with AD. Others have hundreds of different known mutations, and presumably many more yet to be discovered. BRCA1, BRCA2 and Presenilin-1 are examples. Many mutations have been observed in only one family. They *might* all have similar effects, for example if they are all extreme protein-truncating mutations, and early analysis usually assumes that they do, but then again they might not. This heterogeneity even among single gene disorders means that the simple gene $\rightarrow$ mutation $\rightarrow$ disease model of monogenic disorders is becoming less and less tenable (Weatherall, 2000). It might be said that our increasing knowledge is currently turning simple diseases into complex ones.

(c) *Multiple genes, same endpoint.* Some diseases may be caused by mutations in more than one gene; BC (BRCA1 and BRCA2), EOAD (Presenilin-1, Presenilin-2 and APP) and adult polycystic kidney disease (APKD, the APKD1 and APKD2 genes) are examples. For rare disorders, with no sporadic occurrences, the knowledge that any family member has a mutation in one of the candidate genes usually allows the presence of that mutation to be inferred from occurrence of the disorder in any other family member.

(d) *Incomplete penetrance.* Few mutations are thought to be 100% penetrant. Even BRCA1, for example, regarded as causing one of the most severe single-gene disorders, has a lifetime penetrance estimated at between 40% and 80%. The lower the penetrance, or the greater its apparent variability across different populations, the harder the gene-disease link is to study, because the kind of censoring illustrated by Equation (1) becomes more common and inescapable.

(e) *Interactions between genes, other genes and the environment.* Genes may play a part in many complex diseases — much heart disease and all cancers, for example — but with complex interactions among specific alleles of many genes, and the environment and lifestyle. Such multifactorial disorders are difficult to study; the main hope lies in very large scale genetic databases, but even then it may be doubted that the numbers will be sufficient to identify the contributions of individual genes to modest extra risks. And, it is increasingly the case that the rôle of other genes and the environment is not confined to multifactorial disorders. As in (b) above, better knowledge of some single-gene disorders shows other genes and environment play important parts in the onset of disease. Just because the *genotype* shows simple Mendelian inheritance does not mean that the *phenotype* need do so.

(f) *Incomplete ascertainment.* A key question is how a particular family was selected for a study? Did it show a pattern of disease unusual enough to be spotted, when a family with one or two fewer cases would have been missed? Very often there is some such reason for recruiting a family to a study, and the more severe the effect of a mutation (remembering that a gene may have many mutations), the more likely it is to be studied. This is called 'ascertainment bias'. An extreme example is BRCA1; much epidemiology was based on the families ascertained by members of the Breast Cancer Linkage Consortium, all on the basis of multiple affected members in several generations. Initial penetrance estimates of about 80% (for breast cancer) have subsequently been revised down to about 40%. Even once initial ascertainment has taken place, many subjects (persons, sibships or families) may be removed from the study because of missing or suspect information (not uncommonly, paternity other than as stated) which can introduce other forms of selection. Many genetic epidemiologists would put ascertainment at the top of their list of problems.

(g) *Coverage and reliability of genetic tests.* 'Coverage' does not here refer to the completeness of testing within families, but to the ability of a test to detect any and every mutation within a gene. This is only possible if the entire gene sequence is examined, which is the most expensive kind of test. Many genetic tests target specific parts of a gene's coding sequence known to contain mutations. This is very reasonable in a research setting, with affected families, but in a population setting it means that a null test need not imply absolutely normal gene function. It is also not true that sequencing a gene leads straight to a useable DNA-based test. It took several years from sequencing the APKD1 gene to developing a reliable test, because it shared so much of its code with other regions on the same chromosome. No doubt the coverage of genetic tests will improve rapidly.

(h) *Geographical diversity.* Allelic frequencies differ widely across populations, defined by ethnic or geographical origin. Obvious examples are the sickle cell and thallasaemia mutations, that evolved in malarial regions because a single mutated allele gives pro-

tection, although two mutated alleles lead to disease. Combining different population groups, whether in original analyses or meta-analyses, is one way to obtain sizeable study populations, and it may be unavoidable in beginning to understand rare mutations, but it introduces an element of uncontrolled heterogeneity that might be hard to untangle later on.

(i) *Timescale.* Genetic epidemiology just cannot be done at the same pace as molecular genetics. Consider the stages that might be involved in obtaining reliable age-related rates of onset:
   (1) A genetic locus is associated with a disease, by statistical linkage analysis. (This stage might soon be redundant, but in the past it has absorbed huge effort.)
   (2) The gene is found and sequenced, and eventually genetic tests are developed.
   (3) The function of the gene product is discovered, and its rôle in normal metabolism and disease is found.
   (4) Rates of onset are estimated by retrospective studies of affected families, with a high degree of ascertainment bias.
   (5) Prospective population based studies are carried out, if the disorder makes it medically and commercially worthwhile.

   The first three of these stages are molecular biology, only the last two are epidemiology. Any of these stages can take years, none of them is automatic. All in all, the timescale might be measured in decades rather than years. From an insurance point of view this is very bad, because *concerns* about insurance spring forth fully-formed as soon as the gene-disease link is made, but the means to evaluate the true risk become available much nearer the end of this process.

(j) *Terminology.* The investigator who would cross disciplines must be alert to semantic differences. The word 'predictive' is an example: clinical geneticists regard as 'predictive', information that changes the probabilities of disease (maybe small to begin with) by so much that it would change the advice they would give to an individual patient. Actuaries regard as 'predictive', information that changes the expected numbers of claims among an insured population; a much lower threshold. A 50% increase in a 1% probability of disease might not be predictive in a clinical setting, but often would be in an insurance setting.

3.5 *What Do Genetic Epidemiologists Publish?*

Since we must so often rely on published information, we shall give a few examples of what to expect. Of course, it is always possible to contact authors and ask for more data, and some authors are indeed very helpful in this respect. It may be worth noting, however, that some of the following examples describe the information we had *after* receiving such assistance. Original data are almost never available.

(a) *APOE.* The APOE gene was discovered at an early stage in molecular genetics; the gene product was attributed to a single genetic locus with three common alleles $\varepsilon 2$, $\varepsilon 3$ and $\varepsilon 4$ as early as 1981 (Zannis, Just & Breslow, 1981). Also, the gene was discovered *via* the gene product, not the other way round, so the rôle of that gene product has also been studied for an unusually long time, and it was implicated in AD in 1991 (Pericak-Vance *et al.*, 1991). So, by the standards of other genes, its epidemiology is

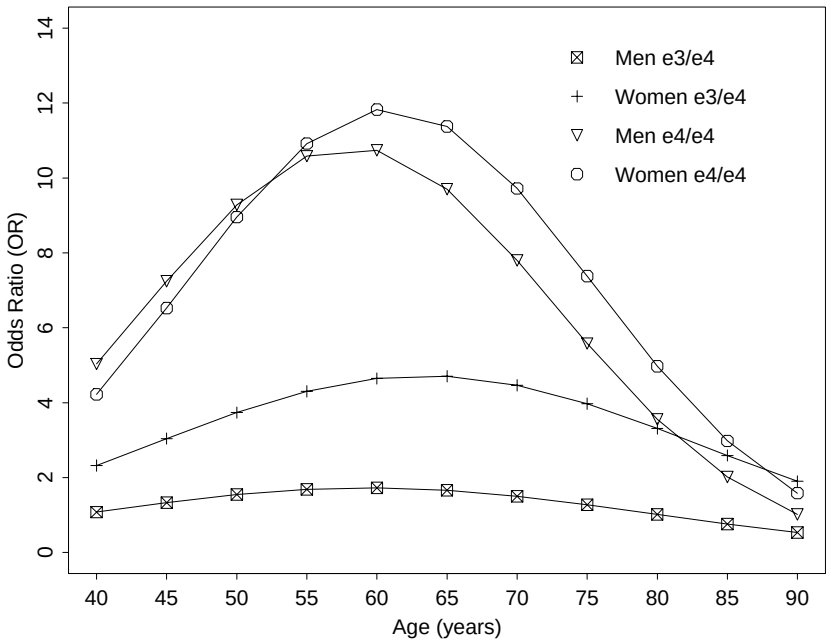

Figure 3: Odds ratios of AD relative to $\varepsilon 3/\varepsilon 3$ genotype for $\varepsilon 3/\varepsilon 4$ and $\varepsilon 4/\varepsilon 4$ genotypes. Source: Farrer *et al.* (1997).

advanced. Macdonald & Pritchard (2000) surveyed the extensive literature and found one study, a meta-analysis by Farrer *et al.* (1997) which reported useable age-related risks of AD by genotype. The information was in the form of odds ratios with respect to the most common genotype, and an example is given in Figure 3. No confidence intervals were published, and in this case the authors kindly provided us with the exact numbers illustrated in the graph.

This study was derived from case-based studies, so significant ascertainment bias is to be expected. After discussions with epidemiologists, the excess rates of onset associated with the $\varepsilon 4$ allele were reduced by 50% and 25% for use in actuarial models.

(b) *BRCA1 and BRCA2.* Ford *et al.* (1998) published age-related penetrance estimates of the onset of BC and/or OC attributable to BRCA1 and BRCA2 mutations. In each case, the reported results were point estimates at ages 30, 40, 50, 60 and 70, with confidence intervals. Other studies have reported only the total penetrance by some high age, giving support for reductions in the rates of onset similar to those used for APOE above.

(c) *APKD.* Several studies have reported survival rates free of end-stage renal disease (ESRD, meaning kidney failure) in respect of APKD1 and APKD2 mutation carriers, in the form of Kaplan-Meier estimates. These are always reported as graphs, often small graphs, without confidence limits. In one case the authors provided us with the underlying figures, but otherwise the only possible approach is to read the figures from the graphs. In this we are not alone: Tudur *et al.* (2001), writing about meta-

analysis, said:

> "The general quality of the actuarial and Kaplan-Meier curves varied across trials and extracting the relevant information from these survival curves can only be approximate. The main problems related to the size and scale of plots and the thickness of curves. Enlarging the graphs by using a photocopier may improve the accuracy."

(d) *Presenilin-1.* As mentioned before, there is almost no epidemiology of EOAD. However, because almost all known mutations in Presenilin-1 are point mutations, there are many papers describing their discovery, some of which include pedigrees. Gui & Macdonald (2002) surveyed this literature (the references alone in this paper occupied nearly nine pages) and used these pedigrees to estimate rates of onset of EOAD. The literature on Presenilin-2 and Amyloid Precursor Protein (APP) mutations, which also cause EOAD, is of a similar nature but less extensive, and this approach was fruitless; until such time as epidemiologists with access to source data choose to study this problem, it will be impossible to say much that is quantitative about these two genes.

4. The Actuarial Toolkit

To date, most studies that have successfully imported genetic epidemiology into actuarial problems have used multiple-state models, and that is the approach we summarise here. States and transitions can be chosen to represent many useful features of actuarial and genetical problems; in particular the discreteness of the states corresponds well with the small number of discrete genotypes that feature in Mendelian single-gene disorders. In this sense, multiple-state models of heterogeneous populations may be viewed as frailty models with a discrete frailty. Classical frailty models (Hougaard, 2000) might be more suitable for modelling multifactorial disorders, but that lies in the future.

Multiple-state models should by now be reasonably familiar, but it appears that many actuaries may be dissuaded from using them because of perceived difficulties, such as solving differential equations. Therefore a short summary, illustrated by some genetical applications, is worthwhile.

4.1 *Transition Intensities and Occupancy Probabilities*

Suppose we have a set $\mathcal{S}$ of $M+1$ states: $\mathcal{S} = \{0, 1, \ldots, M\}$. Transition between states j and k at age $x+t$ is governed by the age-dependent intensity μ_{x+t}^{jk}, interpreted as:

$$\mathrm{P}[\text{Transition } j \to k \text{ by age } x+t+dt \mid \text{In state } j \text{ at age } x+t] = \mu_{x+t}^{jk} dt + o(dt).$$

Define ${}_tp_x^{jk}$ to be the probability that a person in state j at age x shall be in state k at age $x+t$. These occupancy probabilities are related to the intensities by the Kolmogorov forward equations. For all j and k:

$$\frac{\partial}{\partial t} {}_tp_x^{jk} = \sum_{l \neq k} {}_tp_x^{jl} \mu_{x+t}^{lk} - \sum_{l \neq k} {}_tp_x^{jk} \mu_{x+t}^{kl}.$$

Only in special cases can these be solved explicitly, but that is no problem, as they are just linear ordinary differential equations, easily soluble numerically. Standard methods can be found in any textbook on numerical analysis, such as Conte & de Boor (1972), and many standard mathematics packages can be used. After all, the ordinary life table is nothing more than the numerical solution of the Kolmogorov equation:

$$\frac{\partial}{\partial t}{}_t p_x = -{}_t p_x \mu_{x+t}. \tag{2}$$

Intensities are exactly the same as the rates of onset or incidence rates obtained by epidemiologists. More popular in their literature, though, are penetrance estimates, often expressed as Kaplan-Meier estimates of the 'survival' probabilities to onset of the disorder. These are easily converted into intensities through Equation (2).

It is sometimes believed, wrongly, that only multiple-state models with constant intensities can be used. Constant intensities may help in two ways:

(a) If explicit expressions for occupancy probabilities are required, simple intensities are needed, and constant intensities are simplest. However, since many if not most actuarial problems can be handled by solving Thiele's equations (see Section 4.2) with suitable payments, and these do not involve the occupancy probabilities, focussing on the latter might be simply a throwback to the clerical convenience of a discrete-time life table.

(b) When estimating intensities, it is often helpful to assume they are constant over short age or time intervals, because then simple occurrence/exposure rates are obtained:

$$\hat{\mu}^{jk} = \frac{\text{No. of transitions } j \to k}{\text{Total time spent in state } j}.$$

This is purely to help in estimation, and the resulting crude estimates are usually smoothed for practical use. Alternatively, non-parametric methods such as Nelson-Aalen estimates of the integrated intensities can be used (see Collett (1994) for a simple introduction, or Andersen *et al.* (1993) for a full treatment).

4.2 *Payments and Policy Values*

Payments of two kinds are easily attached to a multiple-state model; we assume that positive cashflows are income to the insurer:

(a) a continuous annuity of rate b^j_{x+t} per annum, while in state j at age $x+t$; and

(b) a lump-sum payment of b^{jk}_{x+t} on transition from state j to state k at age $x+t$.

For simplicity, suppose there is a constant force of interest δ. Then the expected present value (EPV) of the future discounted loss (net outgo minus income), at age $x+t$, given presence in state j, is denoted ${}_tV^j_x$ and is the usual prospective policy value. The policy values in all states may be computed as the solution of Thiele's differential equations (Hoem, 1969):

$$\frac{\partial}{\partial t}{}_tV^j_x = \delta_t V^j_x + b^j_{x+t} - \sum_{k \neq j} \mu^{jk}_{x+t} \left(b^{jk}_{x+t} + {}_tV^k_x - {}_tV^j_x \right)$$

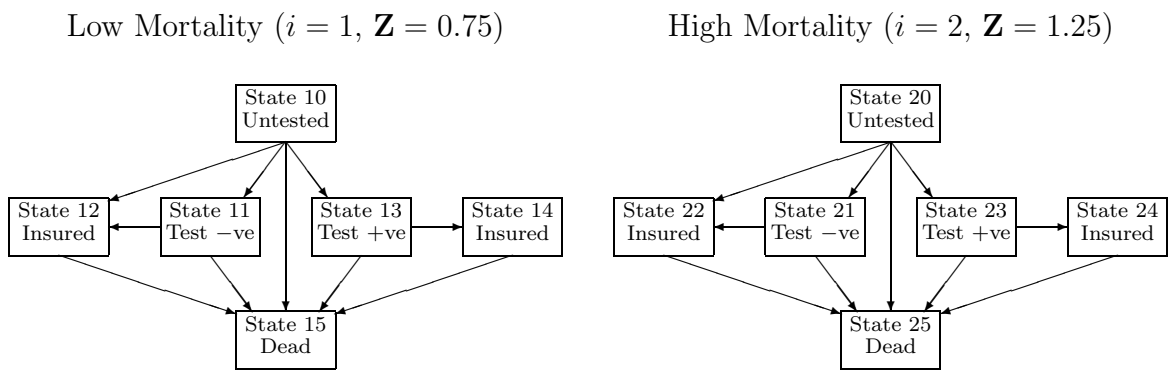

Figure 4: A Markov model of the Ordinary Rates (OR) class with two sub-populations. $\mathbf{Z}$ is a 'frailty' random variable; a person chosen at random has 75% or 125% of 'average' mortality, each with probability 1/2, representing an OR class extending to about 150% of average mortality. Based on Macdonald (1999).

solving backwards from the terminal policy values in each state.

4.3 *Applications*

Two approaches have appeared in the actuarial literature. The first (Macdonald 1997, 1999, 2001) is a top-down approach in which a simple model of all genetic disorders (either single-gene or multifactorial) is used, with no attempt to model the effects of particular genes. Figure 4 (from Macdonald (1999)) shows an example: the Ordinary Rates (OR) class in life insurance (those charged the standard rate of premium) is represented by low- and high-mortality subgroups, and it is assumed that all multifactorial disorders appear in the latter. Various assumptions about the frequency of such disorders, and genetic testing that can detect them, can be made and the costs of adverse selection can then be found. Such models are useful in search of 'null results' along the lines of 'even under unfeasibly extreme (adverse) assumptions, the cost of adverse selection would be negligible'. Their advantage is that they need only the broadest guidance from genetic epidemiology, so they are not held up awaiting progress from that field.

In circumstances where either premium loadings or adverse selection might not be negligible, the top-down approach is less appropriate; a bottom-up approach based on realistic models of particular disorders is needed. Given the current state of genetic epidemiology, this is more a program for the future than a project to be completed now.

We illustrate some of the methodologies that can be employed to answer actuarial questions about genetics, and even perhaps to make an actuarial contribution to genetic epidemiology.

(a) Figure 5 illustrates a multiple-state model of a heterogeneous population, in this instance with respect to APOE genotype. The idea of using multiple-state models to represent such heterogeneity goes back at least as far as Hoem (1988). For long-term care (LTC) insurance modelling, where it may be reasonable to suppose that insurance is paid for by a single premium at an advanced age, we may assume that applicants for insurance are initially healthy, so are distributed over the states $10, 20, 30, 40, 50$ and

60 according to the population genotype frequencies shown. Then adverse selection may be simply represented by increasing the proportion of applicants with riskier APOE genotypes, $\varepsilon 4/\varepsilon 4$, $\varepsilon 3/\varepsilon 4$ and possibly $\varepsilon 2/\varepsilon 4$. Macdonald & Pritchard (2001) did this (allowing for the fact that AD does not account for all LTC costs) and found that only in very small LTC insurance markets, with strong adverse selection and relative risks unadjusted for ascertainment bias (see Section 3.4) were the costs of adverse selection significant. Moreover, only under the same circumstances would the theoretical extra premiums associated with the riskier genotypes be high enough to attract an extra premium in practice.

This kind of 'null result' appears quite often in the conclusions of genetics and insurance modelling, either because the genes concerned are so rare that adverse selection appears negligible, or sometimes (as in the case of APOE) because the risks associated with certain genotypes might not, in fact, be so great compared with more familiar underwriting risks.

Clearly, such a model can be adapted to represent critical illness (CI) or life insurance. The latter usually requires onset and death after onset to be modelled as separate transitions, so CI insurance is often the easier to model in the context of genetics, although it is much less studied otherwise.

(b) Adverse selection can be modelled by adding states to represent uninsured and insured lives. Figure 6 (from Macdonald, Waters & Wekwete (2001b)) shows a model of a CI insurance *market* (not just a single policy or homogeneous cohort of policyholders) in which people who are initially uninsured and who have not had a genetic test may choose to buy insurance before or after being tested. The model represents a given genotype, so the rates of CI events and of death are the same for states $i0$ to $i3$. The population is represented by having one such model for each genotype, along the lines of Figure 5, with appropriate genotype frequencies. By suitable choices of intensities we can represent:

(1) mutation or genotype frequencies;
(2) the prevalence of genetic testing;
(3) the size of the CI insurance market (by the 'normal' rate of purchase);
(4) the extent of adverse selection, meaning both the rate of insurance purchase and the amount purchased; and
(5) the premiums charged, given the insurer's knowledge.

Calibrating such a model is an interesting challenge. The problems of translating genetic epidemiology into rates of onset were mentioned in Section 3, but in addition there is no generally accepted model of CI insurance, and no published insured lives data. The closest published approaches are perhaps Dinani *et al.* (2000), Macdonald, Waters & Wekwete (2001b) and Gutiérrez & Macdonald (2001), all mainly based on population statistics and therefore not necessarily suitable for use without adjustment. However, the questions that arise in connection with genetics are usually about *relative* costs — additional premiums for at-risk applicants or increases in premium rates to pay for adverse selection — so this drawback has less force.

Assuming the equivalence principle is used within each underwriting class in the absence of adverse selection, we have:

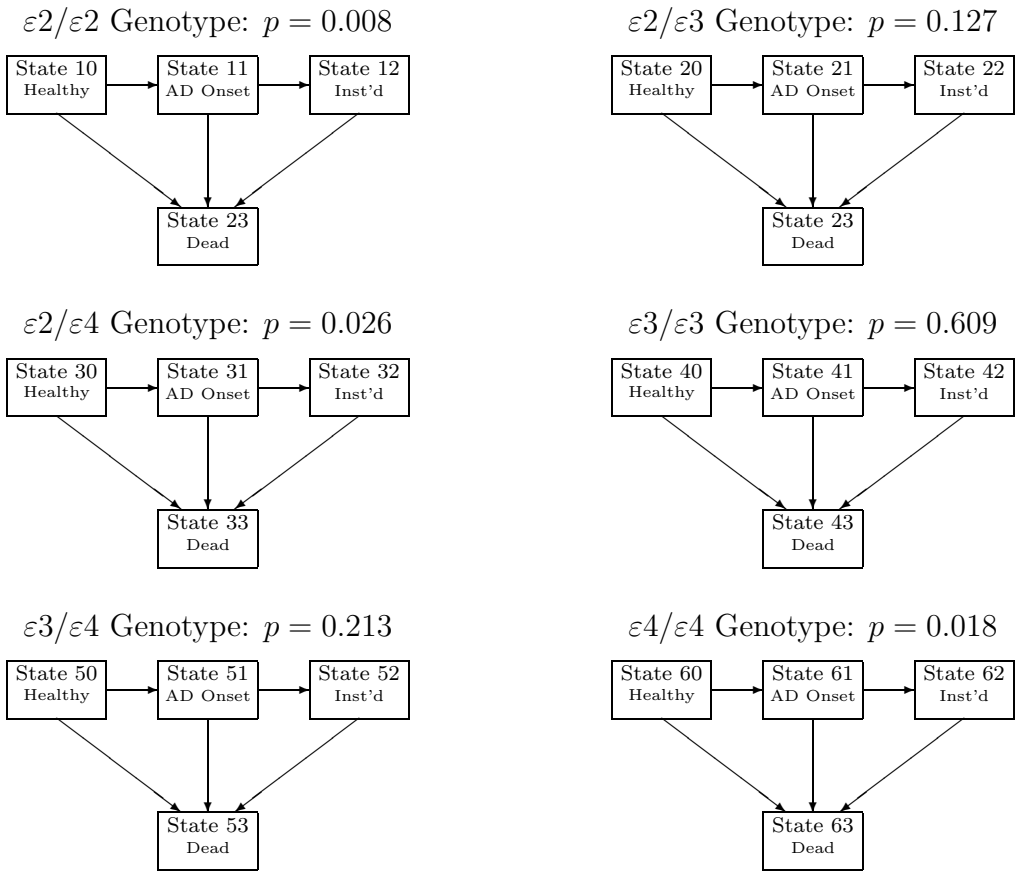

Figure 5: A multiple-state model of Alzheimer's disease in the population. There are six sub-populations defined by APOE genotype, with relative frequencies based on Farrer *et al.* (1997). These define the initial distribution across states 10, 20, 30, 40, 50 and 60. In practice the $\varepsilon 2/\varepsilon 2$ and $\varepsilon 2/\varepsilon 3$ genotypes may be combined because the former is so rare; see Macdonald & Pritchard (2000).

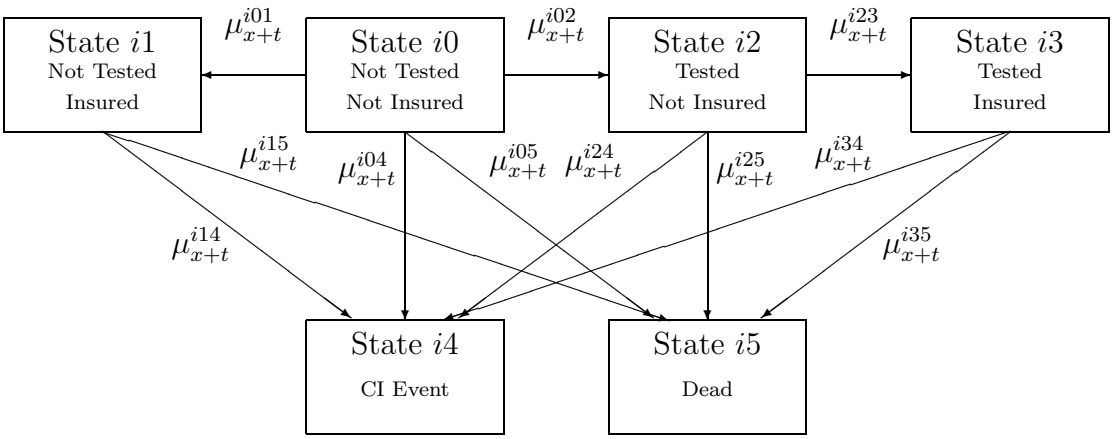

Figure 6: A Markov model of the insurance purchase and CI insurance events for a person with the ith of several genotypes. Source: Macdonald, Waters & Wekwete (2001b).

$$\text{EPV[Insurance loss without adverse selection]} = 0.$$

Then a simple measure of the cost of adverse selection is:

$$\frac{\text{EPV[Insurance loss with adverse selection]}}{\text{EPV[Premiums payable]}} - 1$$

which shows by how much premiums would have to increase to absorb the cost.

To model regular premium contracts, with continuous payments for simplicity, it is necessary that the rate of premium in each underwriting class is the weighted average of the intensities from the healthy states in that underwriting class into the claiming states. Otherwise, if premiums were level, they would depend on the age at purchase of insurance, and would not be adapted to the Markov process framework. In computational terms, this means solving the Kolmogorov equations forward, using the occupancy probabilities to weight the intensities, then using the resulting rates of premium to solve Thiele's equations backwards.

The first such 'bottom-up' study based on particular genes was that into life insurance by Lemaire *et al.* (2000) and Subramanian *et al.* (1999), in respect of BC, OC and the BRCA1 and BRCA2 genes. Broadly, they concluded that the additional knowledge gained from genetic tests was less important than that contained in the family history.

(c) States in a multiple-state model may be grouped into underwriting classes, within each of which a common premium rate can be charged, leading to models of various moratoria on the use of genetic information. In particular, for rare Mendelian disorders which have no cause other than mutations in known genes (such as APKD, HD or EOAD) family histories can be simply represented. Figure 7 (from Gutiérrez & Macdonald (2001)) shows a CI insurance market in a population with three subgroups:

(1) $i = 1$: those not at risk of APKD, because they have no known family history;
(2) $i = 2$: those at risk of APKD because they have a family history, but who are *not* mutation carriers; and
(3) $i = 3$: those at risk of APKD because they have a family history, who *are* mutation carriers.

By Mendel's laws, and given that APKD mutations occur in about 1 per 1,000 of the population, we can assign the proportions 0.998, 0.001 and 0.001 respectively to these three groups, from birth up to the lowest age when APKD appears (thereafter, the relative proportions must be found by solving the Kolmogorov equations).

Macdonald (2001) and Gutiérrez & Macdonald (2001) modelled the effect of moratoria on all genetic test results, on adverse test results only, and on a family history of a Mendelian disorder. The former was a 'top-down' study of mortgage-related life insurance, the latter a 'bottom-up' study of APKD and CI insurance. Both found that extending a moratorium to family history may cause premiums to rise simply because persons at higher risk are offered standard rates; the costs of adverse selection in addition (if these persons did more than simply adopt 'normal' insurance-buying habits) could be significant in a small market.

(d) Diseases which commonly occur for reasons other than mutations in known genes (such as BC and OC) are simple to model assuming a known mutation, but much

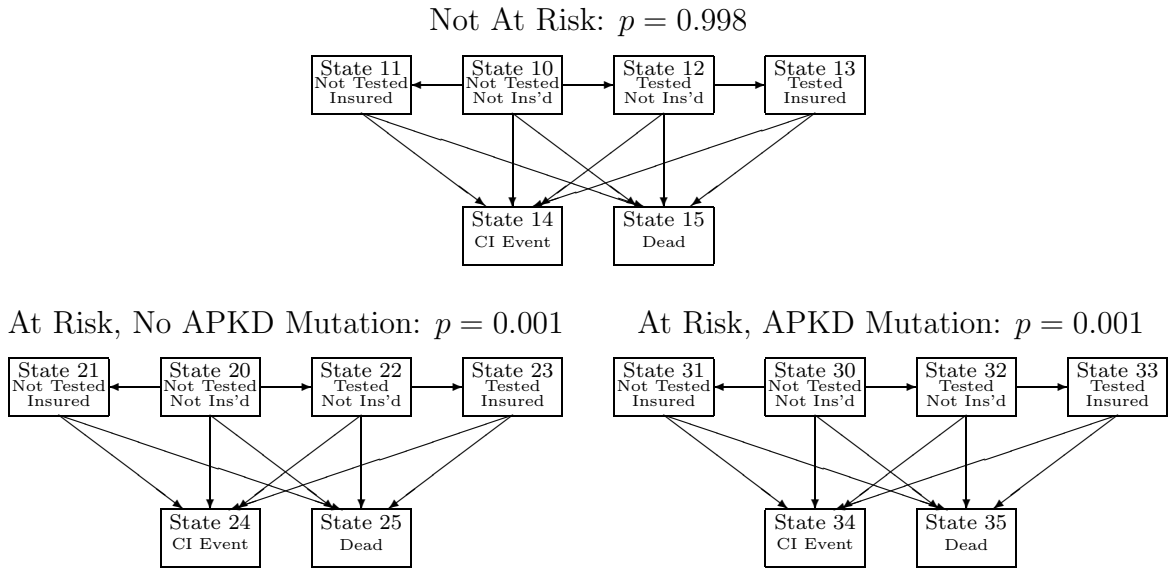

Figure 7: A Markov model of a Critical Illness insurance market allowing for family history of APKD. Source: Gutiérrez & Macdonald (2001).

more difficult to model assuming only family history is known (as would be common in underwriting). Assuming that the rates of onset $\mu_g^{BC}(x)$ and $\mu_g^{OC}(x)$ can be found for each genotype g, it is simple to compute:

$$\text{P[Onset by age } x \mid \text{Genotype } g]$$

for one person. From this, we proceed through a series of increasingly demanding applications of Bayes' Theorem, illustrated by Macdonald, Waters & Wekwete (2001a) in the case of BC and OC.

(1) Given the number M of female relatives of the applicant (including her as the first), the vector of their ages $X = (x_1, x_2, \ldots, x_M)$ when the applicant was born (so $x_1 = 0$), and the vector of all their genotypes $G = (g_1, g_2, \ldots, g_M)$, we can compute the probability of the family history observed when the applicant has reached age x, defined as the vector of states occupied by every family member:

$$\text{P[Family history } |M, X, G].$$

(2) Assuming M and X to be known, we can find:

$$\text{P}[G|M, X \text{ and family history}].$$

(3) All we are interested in is the genotype of the applicant, g_1. Integrating out all the other genotype information we get:

$$\text{P}[g_1|M, X \text{ and family history}].$$

(4) The underwriter will not usually know the histories of all the female relatives. For example, the number of affected aunts could be unknown. Integrating out the unknown information again, we find:

$$\mathrm{P}[g_1|M, X \text{ and known subset of family history}].$$

(5) Finally, M and X might be unknown; underwriters usually do not know anything about unaffected relatives. Given estimates of the joint distribution of M and X, we compute:

$$\mathrm{P}[g_1 \mid \text{Known subset of family history}]$$

which corresponds most closely to underwriting practice.

Once these conditional probabilities are found, they can be used with a model of underwriting classes along the lines of Figure 7 to study the effects of various moratoria on using genetic information. The extra complexity introduced by sporadic cases of disease, not caused by the gene(s) of interest, is clear.

The number of calculations involved is proportional to:

$$(\text{Number of distinct genotypes})^M$$

and even after making some assumptions that reduced the number of genotypes from 9 to 4, these computations were at the limit of the computing power available for the study cited above.

(e) States may represent events that reveal information needed for estimation, as well as events related to insurance and insurance purchase. Consider the basic problem of family history data mentioned at the start of Section 3.4: unless a person has a genetic test or develops a dominant disorder, we may never know whether or not they were a mutation carrier. Therefore the individual exposures in occurrence/exposure rates must be weighted in some way by the probability that an unaffected at-risk person is a mutation carrier, which at any age x depends on:

(a) whether or not they have had a (reliable) genetic test; and

(b) which of their siblings, parents, grandparents and so on are known carriers.

The three states on the left-hand side of Figure 8 represent the events befalling an individual who may start in state 0 (Presenilin-1 mutation carrier) or state 1 (non-carrier) but which may be unknown. The *probability* that they were born into state 0 changes as information is acquired from events in the three states on the right-hand side, which also involve the person's relatives (see Gui & Macdonald (2002)). The model leads to estimates of the intensity of interest, μ_x^{02}, the rate of onset of EOAD among known mutation carriers. The same intensity μ_x^{02} may be involved in the intensities ν_x^{jk} in the right-hand (information) part of the model, because onset in relatives may be the cause of a transition there, but this specification in terms of a multiple-state model (strictly, a multivariate counting process) tells us precisely what kind of estimates we get for μ_x^{02} if we use events in the three right-hand states only for conditioning.

This model embodies the principle that estimates of rates of onset at age x are conditioned only on information known at age x, here represented by the three right-hand states. This principle seems not to be applied uniformly in genetic epidemiology,

Figure 8: A multiple-state model of early-onset Alzheimer's disease incorporating information about risk of being a mutation carrier (see Gui & Macdonald (2002)). $\otimes$ indicates that the state space is $\mathcal{S} \times \mathcal{P}$, where $\mathcal{S}$ is the three states on the left and $\mathcal{P}$ is the three states on the right. The problem is to estimate μ_x^{02}, the rate of onset of EOAD among known carriers of Presenilin-1 mutations. The conditional probability of a healthy person being a mutation carrier (state 0 in the left-hand set of states) depends on the state occupied in the right-hand set of states, which represents what is known about relatives and/or genetic test results.

where it may be common to condition on known genotypes, even if it was eventual onset of the disease that revealed the genotype. Newcombe (1981) is an example. That this can be problematic can be seen by noting that in a formally specified probability model, the acquisition of information is represented by observation of events; we should not condition on 'knowing' the genotype, but on observing the event that revealed it. Then it is clear that attempts to base estimation on conditional probabilities:

$$\mathrm{P}[\text{Onset at age } x \mid \text{Mutation carrier}]$$

should properly be based on conditional probabilities:

$$\mathrm{P}[\text{Onset at age } x \mid \text{Onset observed at age } x + t]$$

which are identically 0 or 1, and the attempt fails. Similar logic applies to knowledge gained from events other than onset in the individual concerned. This is an interesting area for further research where actuarial science and genetic epidemiology overlap.

4.4 *Summary*

It has been shown that multiple-state models are capable, in principle, of answering many important questions relating to small numbers of genes with medium to high penetrance, including:

(a) estimation of extra premiums;
(b) estimation of the possible costs of adverse selection;
(c) possible allowances for ascertainment bias; and
(d) the impact of different forms of moratoria.

'Top-down' models are useful for seeking 'null' results, where even extreme assumptions lead to no serious consequences. In the longer term, 'bottom-up' models based on

good estimates of mutation frequencies and penetrances are needed. The limiting factor just now is the need for more progress in genetic epidemiology. Computing power could also impose limits, but no very great computing power has yet been applied, and other methods (for example, simulation) might prove useful. Different methods altogether might be needed to model multifactorial disorders.

5. Evidence-Based Underwriting and Policy-Making?

5.1 *What Basis for Legislation or Regulation?*

The U.K. is relatively unusual, because the government has consistently declined to use its powers to intervene in the matter of genetics and insurance. Instead, it has appointed commissions to advise it (the Human Genetics Advisory Commission (HGAC) in 1996, subsumed into the Human Genetics Commission (HGC) in 2000). Their approach has been to seek evidence upon which to base their advice. The default models in existing legislation are perhaps the Acts regulating discrimination on the grounds of sex and disability. These outlaw almost all discrimination, but exempt insurers if there is actuarial or statistical evidence of a genuine difference in insurance costs. Very few test cases have appeared in courts. Another important commentator has been the House of Commons Science and Technology Committee (HCSTC), a committee of parliament free to investigate and to comment upon governent policy. The history so far is as follows.

(a) An HCSTC report (HCSTC, 1995), on all aspects of human genetics, followed a period of stalemate, during which the ABI had stood firm on its 'right to underwrite', and had won few friends. The report did not favour the industry, and advised giving it a year to find some accommodation with genetics issues, advice which the government rejected, instead forming the HGAC and asking it for recommendations.

(b) The HGAC report (HGAC, 1997) recommended a two-year moratorium on the use of genetic test information, essentially as a breathing space while research could be done. It was particularly pointed on the industry's lack of research.

(c) Meanwhile, the ABI had appointed a clinical geneticist as its genetics advisor, and had introduced its voluntary moratorium:
   (1) banning insurers from asking anyone to take a genetic test;
   (2) ignoring any DNA-based test results for life insurance of up to £100,000 in connection with a mortgage; and
   (3) forbidding 'cherry-picking' (using favourable genetic test results to offer lower premiums than standard).

   The ABI also introduced a list of eight, later seven, single-gene disorders it regarded as significant for insurance. These were all disorders which would result in a Mendelian family history; in fact the only reason why anyone might be tested for them would be having a strong family history.

   Perhaps influenced by these actions, the government rejected the HGAC's recommendation for a wider moratorium. In effect, this meant that the ABI's definition of genetic information — the narrowest possible, based on direct examination of DNA or chromosomes — was tacitly adopted.

(d) The government did, however, set up the Genetics and Insurance Committee (GAIC) with the task of considering applications from insurers (in reality, the ABI) to be

allowed to use specific genetic test results in connection with specific forms of insurance. GAIC decided that 'significant' meant additional mortality of +50% or additional morbidity of +25%, and devised a form indicating what medical and actuarial evidence it expected to see. In late 2000 it approved an application in respect of HD and life insurance, which had largely been based upon Smith (1998).

(e) Surprisingly, the ABI advised its member companies that they could continue to use test results until told not to by GAIC, rather than suspending their use pending approval. This was enormously controversial (and quite unnecessary from a financial point of view, since family history was still allowed in underwriting).

(f) Two reports were issued in 2001 (HCSTC, 2001; HGC, 2001), both extremely critical of the industry. The significant consequences were as follows.
   (1) The HGC queried the narrow focus on DNA-based genetic tests, and indicated that it regarded genetic information more broadly (see Zimmern (2000) for a discussion of genetic information). In particular, the HGC regarded family history as genetic information.
   (2) A new five-year moratorium was agreed between the government and the ABI; genetic test results would be ignored for life insurance up to £500,000, and for other insurances up to £300,000.
   (3) Family history was still not covered by the moratorium, but the HGC said that after three years it would expect insurers to produce the evidence upon which they based family history underwriting.

In summary, the U.K. government, through its commissions, has been willing to take an evidence-based approach to policy-making. In return, the industry is being asked to show that it takes an evidence-based approach to underwriting.

5.2 *Normal Science* versus *Actuarial Science*

'Normal' science has a 'gold standard' of peer-reviewed publication in refereed journals. It is not perfect, but it is the standard by which others will be judged. In the U.K., HGAC, HGC and GAIC have gathered evidence from medical science and from actuarial science, and set them side-by-side. How do they compare? The answer may be: not too well.

(a) Actuarial science has not more than four international journals whose purposes, procedures and reputations fully match those of 'normal' science (ASTIN Bulletin, Insurance: Mathematics & Economics, North American Actuarial Journal and Scandinavian Actuarial Journal).

(b) Academic actuaries, who are more accustomed to the standards set by peer-reviewed publication, are few in number. Actuaries in commercial practice are often used to a looser framework of validation: argument from experience or authority; informal review by chosen colleagues; and 'publication' in company papers or by presentation to meetings. Commercial confidentiality can also play a part. For many purposes, of course, that is sufficient, but all the conveniences of this less formal approach contain also the very sins that peer-reviewed publication has evolved to guard against.

(c) With a few notable exceptions, such as the Continuous Mortality Investigation Bureau in the U.K., there is little spirit of co-operation between companies or of supporting

long-term research. This is particularly unfortunate, since national scientific funding bodies are often reluctant to sponsor research impinging on commerce, that it appears that industry could be supporting by itself.

(d) Bodies advising on public interest issues are increasingly accountable and transparent. The evidence they consider is often in the public domain. Experience and authority are not therefore excluded, but their basis is liable to be examined. It should not be surprising if the process of 'normal' science is given much greater weight than the less formal, often invisible processes of commerce.

In the past, insurers have been able to apply a precautionary principle in their own favour: when ignorant about risks, premiums could be set conservatively to protect the insurance pool, or cover denied. This might have been diluted by competition in practice, but there was no other constraint. The effect of the public pressure over genetics could be to reverse that position: the precautionary principle may have to be exercised in favour of the insured. That is the practical effect of the various moratoria in force. A precautionary principle may be relaxed or changed as evidence replaces ignorance — that is part of its purpose — but that supposes that convincing research produces evidence.

It is notable that the public bodies in the U.K. that have examined the issue have kept returning to the extent of the actuarial research base, and clearly have had peer-reviewed publication in mind; Aitken (1998) made this clear. Genetic testing may have brought evidence-based underwriting closer, meaning that practices may be subjected to a degree of proof before use, rather than suffering the occasional legal challenge in retrospect. But family history is not far behind, and unless genetic information continues to be regarded as special, any part of underwriting could follow.

Acknowledgements

This paper is based on talks given in Stockholm in November 2001, at a workshop on actuarial research in genetics and insurance organised by the Svenska Aktuarieföreningen. The author is Director of the Genetics and Insurance Research Centre at Heriot-Watt University, Edinburgh, which receives funding from a number of insurance companies through the Association of British Insurers.

References

Aitken, C. (1998). Should insurers know our genes?. *Transactions of the* 26th *International Congress of Actuaries, Birmingham*, **9**, 109–115.

Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

ABI (1999). *Genetic testing: ABI code of practice (revised August 1999).* Association of British Insurers, London.

Collett, D. (1994). *Modelling survival data in medical research.* Chapman & Hall, London.

Conte, S.D & de Boor, C. (1972). *Elementary numerical analysis.* McGraw-Hill.

Crowder, M.J. (2001). *Classical competing risks.* Chapman & Hall/CRC.

Dartigues, J.F. & Letenneur, L. (2000). Genetic epidemiology of Alzheimer's disease. *Current Opinion in Neurology*, **13**, 385–389.

Davies, K. & White, M. (1995). *Breakthrough. The quest to isolate the gene for hereditary breast cancer.* Macmillan.

Dinani, A., Grimshaw D., Robjohns N., Somerville S., Spry A., Staffurth J. (2000). *A critical review. Report of the Critical Illness Healthcare Study Group.* Unpublished paper, presented to the Staple Inn Actuarial Society.

Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., van Duijn, C.M. & APOE and Alzheimer Disease Meta Analysis Consortium (1997). Effects of age, gender and ethnicity on the association between apolipoprotein E genotype and Alzheimer's disease. *Journal of the American Medical Association*, **278**, 1349–1356.

Fischer, E.-P. & Berberich, K. (1999). *Impact of modern genetics on insurance.* Publications of the Cologne Re, No. 42.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A.J., Gayther, S.A., Birch, J.M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, A., Hamann, U., Haites, N., Scott, R.J., Maugard, C.M., Vasen, H., Seitz, S., Cannon-Albright, L.A., Schofield, A., Zelada-Hedman, M. & the Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *American Journal of Human Genetics*, **62**, 676–689.

Gui, E.H. & Macdonald. A.S. (2002). A Nelson-Aalen estimate of the incidence rates of early-onset Alzheimer's disease associated with the Presenilin-1 gene. *To appear in ASTIN Bulletin.*

Gutiérrez, M.C. & Macdonald, A.S. (2001). Adult polycystic kidney disease and critical illness insurance. *Research Report No. 01/4, Department of Actuarial Mathematics and Statistics, Heriot-Watt University.*

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. (2000). *Analysis of multivariate survival data.* Springer.

HCSTC (1995). *House of Commons Science and Technology Committee, Third Report: Human genetics: the science and its consequences.* H.M.S.O., London.

HCSTC (2001). House of Commons Science and Technology Committee, Fifth Report: Genetics and insurance. *Unpublished manuscript at* www.publications.parliament.uk/pa/cm200001/cmselect/cmsctech/ 174/17402.htm.

HGAC (1997). *The implications of genetic testing for insurance.* Human Genetics Advisory Commission, London.

HGC (2001). The use of genetic information in insurance: Interim recommendations of the Human Genetics Commission. *Unpublished manuscript at* www.hgc.gov.uk/business_ publications_statement_01may.htm.

Lange, K. (1997). *Mathematical and statistical methods for genetic analysis.* Springer.

Lemaire, J., Subramanian, K., Armstrong, K. & Asch, D.A. (2000). Pricing term insurance in the presence of a family history of breast or ovarian cancer. *North American Actuarial Journal*, **4**, 75–87.

Macdonald, A.S. (1997). How will improved forecasts of individual lifetimes affect underwriting?. *Philosophical Transactions of the Royal Society* B, **352**, 1067–1075, and (with discussion) *British Actuarial Journal*, **3**, 1009–1025 and 1044–1058.

Macdonald, A.S. (1999). Modeling the impact of genetics on insurance. *North American Actuarial Journal*, **3:1**, 83–101.

Macdonald, A.S. (2001). Moratoria on the use of genetic tests and family history for mortgage-related life insurance. *Genetics and Insurance Research Centre, Heriot-Watt University: Research Report No. 01/3.*

Macdonald, A.S. & Pritchard, D.J. (2000). A mathematical model of Alzheimer's disease and the APOE gene. *ASTIN Bulletin*, **30**, 69–110.

Macdonald, A.S. & Pritchard, D.J. (2001). Genetics, Alzheimer's disease and long-term care insurance. *North American Actuarial Journal*, **5:2**, 54–78.

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2001a). The genetics of breast and ovarian cancer I: A model of family history. *To appear in Scandinavian Actuarial Journal.*

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2001b). The genetics of breast and ovarian cancer II: A model of critical illness insurance. *To appear in Scandinavian Actuarial Journal.*

Meiser, B. & Dunn, S. (2000). Psychological impact of genetic testing for Huntington's disease: an update of the literature. *J. Neurol. Neurosurg. Psychiatry*, **69**, 574–578.

Neill, A. (1977). *Life contingencies.* Heinemann, London.

Newcombe, R.G. (1981). A life table for onset of Huntington's Chorea. *Annals of Human Genetics*, **45**, 375–385.

Pasternak, J.J. (1999). *An introduction to human molecular genetics.* Fitzgerald Science Press, Bethesda, Maryland.

Pericak-Vance, M.A., Bebout, J.L., Gaskell, P.C., Yamaoka, L.H., Hung, W.Y., Alberts, m.J., Walker, A.P., Bartlett, R.J., Haynes, C.A., Welsh, K.A., Earl, N.L., Heyman, A., Clark, C.M. & Roses, A.D. (1991). Linkage studies in familial Alzheimer's disease: evidence for chromosome 19 linkage. *American Journal of Human Genetics*, **48**, 1034–1050.

Sham, P. (1998). *Statistics in human genetics.* Arnold.

Smith, C. (1998). *Huntington's chorea: A mathematical model for life insurance.* Unpublished manuscript, Swiss Re, Zurich.

Strachan, T. & Read, A.P. (1999). *Human molecular genetics, second edition.* BIOS Scientific Publishers, Oxford.

Subramanian, K., Lemaire, J., Hershey, J.C., Pauly, M.V., Armstrong, K. & Asch, D.A. (2000). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *Journal of Risk and Insurance*, **66**, 531–550.

Tudur, C., Williamson, P.R., Khan, S. & Best, L.Y. (2001). The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *Journal of the Royal Statistical Society A*, **164**, 357–370.

Weatherall, D.J. (2000). Science, medicine and the future. Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases. *British Medical Journal*, **321**, 1117–1120.

Zannis, V.I., Just, P.W. & Breslow, J.L. (1981). Human apolipoprotein E isoprotein subclasses are genetically determined. *American Journal of Human Genetics*, **33**, 11–24.

Zimmern, R. (2001). What is genetic information?. *Genetics Law Monitor*, **1:5**, 9–13.