Genetics, family history, and insurance underwriting: an expensive combination?

Angus S. MacDonald
Department of Actuarial Mathematics and Statistics
Heriot-Watt University, Edimburgh

Jean Lemaire
Wharton School, University of Pennsylvania

Abstract

Since the mid-1990s, genetics has been the focus of arguments about discrimination and insurance that have surpassed even those about gender and disability in their emotional content. Since then, actuarial research has begun to add quantitative insights about the financial implications of the new genetics, but it is still some distance behind the basic science, and there are still important economic and empirical questions to answer. More recently, the debate has embraced the use of family history and, by extension, other medical evidence used in underwriting. We will give a broad review of the problems, the areas where actuaries have made progress, and the challenges for future research.

Outline

1. Legislation concerning the use of genetic tests by insurers
2. Situation in the UK and Australia
3. General features of genetic epidemiology
4. Markov models
5. An example: the BRCA mutation
6. New models and questions
7. A program of future research

1. Legislation concerning the use of genetic tests by insurers

The human genome project was completed in April 2003. Researchers hope that the sequencing of the human genome will allow them to develop new drugs and therapies, and to identify genetic risk factors for a variety of conditions. At the same time, many fear that the human genome map may open a new frontier for potential discrimination, particularly in insurance. Consumer groups fear that advances in genetic testing may lead to the creation of an underclass of individuals who cannot obtain insurance, and want to ban insurance carriers from accessing genetic test results. Insurance companies point to the risk of adverse selection. With over 1,100 genetic conditions identified and over 800 genetic tests offered, they claim that policyholders may gain a financial advantage, through insurance purchase decisions, from genetic information known to them but not revealed to insurers. They fear that, without a level information playing field, a "death spiral" of increasing premiums and decreasing portfolio sizes may threaten their financial solvency.

Consequently, in many countries, an intense legislative and lobbying activity is taking place that could shape the environment of underwriting in life and health insurance. This activity is reminiscent – and even more emotionally charged - of the debates over access to HIV tests of the 1980s, and over the use of gender in rating of the early 1990s. The focus of the debate is nowadays enlarging, as in several countries the use of family history and, by extension, other medical evidence, is being challenged.

The questions that regulation needs to address are

(i) Should insurers be permitted to reflect in their rates the information provided by genetic tests?

(ii) Should insurers be permitted to require applicants to disclose the results of genetic tests taken prior to the application for insurance?

(iii) Should insurers be permitted to require applicants to take genetic tests prior to consideration of the application?

The answers to these questions have led to four major types of laws, which can be classified as follows[1]. Under a *Laissez Faire* approach, insurers have full freedom to request new tests and the disclosure of existing tests, and to incorporate test results in underwriting and rating. This is practiced in Australia, Canada, Japan, Ireland, Portugal and Spain. Through *Disclosure Duty*, applicants have to disclose the results of existing tests, at the insurer's request, but cannot be required to take additional tests; this approach is used in Germany, New Zealand and the UK. By *Consent Law*, applicants are not required to divulge genetic tests results. If they do, insurers may use this information; this approach exists in Netherlands and Switzerland. In Austria, Belgium, Denmark, France, Italy, and Norway, under *Strict Prohibition*, insurers cannot request genetic tests, cannot require applicants to provide existing tests results, and cannot use any genetic information in underwriting and rating.

In the absence of (or in addition to) legislation, three approaches have been used by insurers' associations. Through a *Voluntary Agreement*, the Swedish Insurance Federation and the state have agreed that no genetic tests can be made a condition for issuance or modification of a life insurance policy. In addition, the results of tests taken prior to the application will not be considered in the risk assessment, unless the sum insured exceeds 15 times an inflation-adjusted base amount ($62,000 in 1999). In Germany, Finland, Greece, the Netherlands, Switzerland, and the UK, insurers have adopted a voluntary *Moratorium* on the use of genetic tests. The moratorium may apply to all life insurance policies (Germany) or to all policies with a sum insured under a given

[1] Legislative activity concerning the regulation of genetics in insurance is proceeding at such a fast pace that some of the information in this introduction may have become obsolete today.

limit (Netherlands, UK). Finally, in Australia and South Africa, insurers' associations have put together *Guidelines* or a *Code of Conduct*.

In the United States, where insurance is regulated at the state level, Wisconsin became in 1991 the first state to pass a law barring insurers from using existing genetic test information and requesting new tests from health insurance applicants. Other states have followed suit, periodically revising their laws to adapt to advances in genetic testing. Currently, 44 states prohibit the use of genetic tests results for health insurance risk classification and pricing; 8 states have similar legislation for life insurance contracts. In many of these states, insurers also cannot request genetic information from family members. Interestingly, 3 other states (Arizona, Vermont and West Virginia) have laws whereby insurers may not use genetic testing results "without actuarial justification". This provision would allow insurers to invest in and use medical studies to calculate the mortality risks associated with a particular test and justify its use for future premium determination. At the federal level, Congress is currently debating the "Genetic Nondiscrimination in Health Insurance and Employment Act" (HR 602/S 318). This bill places restrictions on insurers, banning the use of genetic information in underwriting health insurance; it does not address the use of genetic testing results in life insurance underwriting.

The history of genetic testing legislation in two countries, the United Kingdom and Australia, is worth describing in some detail.

2. Developments in the United Kingdom and Australia

Governments in the United Kingdom and in Australia have declined to legislate in haste, and instead have engaged in discussions with their insurance industries and other interested bodies. Although a great deal of the publicity surrounding these processes has been critical of insurers, the discussion has generally led to the examination of the evidence for and against the use of genetic information by insurers, and therefore of the evidence base for underwriting in general.

The United Kingdom

Up to about 1995, the industry responded to geneticists' warnings about the potential predictive power of genetic tests simply by asserting its 'right to underwrite', regarding genetic information as no different from any other medical information. In 1995 the Select Committee on Science and Technology of the House of Commons (HCSTC) issued a report on genetics (HCSTC, 1995), strongly critical of the industry. Government did not accept its recommendations, but did set up the Human Genetics Advisory Commission (HGAC) to advise it. HGAC made insurance its first priority and reported in 1997 (HGAC, 1997) and in doing so focussed on the question of scientific evidence for underwriting practices. Meanwhile, the industry in the form of the Association of British Insurers (ABI) has appointed a clinical geneticist, Professor J. A. Raeburn, as its adviser, and had drawn up a list of eight single-gene disorders of potential significance for insurance. One of these, Adult Polycystic Kidney Disease (APKD) was soon dropped, on the grounds that it was normally detected by ultrasound and not by a DNA-based test, reflecting the narrow definition of 'genetic test' that the ABI had adopted. In addition, the ABI introduced a code of conduct (ABI, 1999) that banned insurers from asking anyone to have a genetic test, or using existing test results for 'cherry-picking', and introduced a moratorium on the use of any genetic test results for life insurance of up to £100,000 in connection with house purchase.

An important feature of the disorders in the ABI's list is that they were all dominantly inherited late-onset diseases, of which Huntington's disease (HD) is the clearest and best-known example. These were exactly the diseases that had long been known to 'run in

families', and which would normally have been disclosed to an insurer anyway on the basis of a family medical history. In many cases the underwriting decision would have been declinature. For these diseases, genetic testing offered a refinement of existing knowledge, not fundamental new discovery of diseases. Indeed, in the clinical genetics services offered by the National Health Service, testing was and is normally restricted to persons whose family history suggests a strongly inherited disorder, so the only people ever likely to have a genetic test would be those whose family history would have been disclosed to an insurer anyway. The moratorium in the U.K., unlike those in some other countries (Sweden for example) did not extend to the use of family medical history. This became significant in view of the setting up of the Genetics and Insurance Committee (GAIC).

GAIC was set up as a committee of the Department of Health, charged with examining applications made by insurers (in practice, the ABI) to be allowed to use specific genetic test results in connection with specific insurance contracts. It decided that its criteria of 'actuarial relevance' would be +50% additional mortality or +25% additional morbidity, broadly following existing industry practice. In late 2000, GAIC approved an application in respect of HD and life insurance. Predictably, this generated much press comment. More important, however, were the tests that GAIC had *not* yet approved. The ABI had advised its members that they could continue to use these tests, within the terms of the moratorium, until GAIC ruled otherwise, when they should refund any extra premiums. No-one else interpreted the agreement underpinning GAIC in this way, least of all the HCSTC, which returned to the fray with a hard-hitting report (HCSTC, 2001) or the Human Genetics Commission (HGC) into which the HGAC had been subsumed, which followed suit (HGC, 2001, 2002). The ABI's advice seems hard to explain, unless in defence of the pure principle of 'right to underwrite', because of the continued ability of insurers to use family medical history. With hindsight, it would have been better to refrain from using any tests until allowed to do so by GAIC. The outcome was a strengthened moratorium (up to £500,000 of life insurance or £300,000 of other insurances), a reformed GAIC with a stronger remit, and a promise by the HGC to come back in 2004 to examine the evidence for the use of family medical history. At the time of writing, GAIC is revising its criteria of 'actuarial relevance', which may cause it even to revisit its decision on HD.

Australia

While the government in Australia was slower to enter the genetics and insurance debate than was the government in the U.K., when it did so the result was possibly the most thorough and searching examination the subject has yet had. In 2001, the Australian Law Reform Commission (ALRC) was asked to inquire into the protection of human genetic information. It produced an issues paper (ALRC, 2001) then a consultation paper (ALRC 2002) and finally a report (ALRC, 2003), all available at www.alrc.gov.au. The report (ALRC, 2003) is quite simply magisterial, and will surely become a primary reference for legislators in many jurisdictions around the world. Professor Raeburn, the ABI's adviser, described it as "... the most clear and least biased description of genetics and insurance ever produced".

In respect of insurance, the work of the ALRC drew significantly on previous work in the U.K., but it seems clear that much of the suspicion and outright hostility towards the insurance industry that was so apparent in the U.K. was muted or absent in Australia, perhaps reflecting much better working relations between government and the industry there.

The ALRC's primary recommendation was that a Human Genetics Commission of Australia (HGCA) should be set up, with wide ranging powers to advise government and provide expertise on all aspects of human genetics. Superficially the HGCA appears to resemble the HGC in the U.K., and the two should have many responsibilities in common. However, if the Australian government implements the ALRC's recommendations in full, the HGCA would be significantly better resourced and more broadly-based. For example, the ALRC accepted (or at least did not rule out) the need for insurance industry representatives and actuaries to be included in some way in the HGCA; these groups are conspicuously absent from the HGC.

The HGCA would "... have a specific role in ... making recommendations about the suitability of specific genetic tests (and the appropriate analysis and treatment of results) for use by the insurance industry (for example for risk-rating purposes), and by employers (for example for occupational health and safety reasons)" (ALRC, 2003, Recommendation 5-3). However, rather than 'approving' or 'disapproving' tests in the manner of GAIC,

the ALRC suggested that the HGCA's recommendations could be given effect through industry codes of practice (ALRC, 2003, Section 5.104). Recommendation 27-2, which gives effect to this, is worth quoting in full:

> "The Investment and Financial Services Association (IFSA) and the Insurance Council of Australia (ICA) should develop mandatory policies for their members to ensure that, once the HGCA has made a recommendation in relation to the use of a particular genetic test in underwriting, that test is used only in conformity with the recommendation. As a transitional arrangement, insurers should be permitted to continue using genetic tests in underwriting in accordance with industry policies, until such time as the HGCA makes a recommendation in relation to those tests."

In other words, the process that proved so controversial in the U.K. is exactly that which the ALRC is recommending! The difference, of course, is in the level of trust in the industry's policies which the ALRC believes is sustainable, and which government in the U.K. seems to believe is not.

Other significant recommendations (27-5 to 27-8) tighten up the requirement for insurers to justify adverse underwriting decisions, (and these include family medical history), allow appeals (27-9), and require insurers to seek a Public Interest Determination under the Privacy Act 1988 "... in relation to the practice of collecting genetic information from applicants about their genetic relatives for use in underwriting insurance policies in relation to those applicants."

3. General features of genetic epidemiology

There is practically no chance of genetic disorders ever being investigated for insurance purposes, in the way that actuaries have been able to study their own insured populations in the past. Insurers have little or no experience of the rarer disorders, because people at risk have tended to be declined in the past, and because their numbers are anyway so small. Also, the existence of moratoria will mean that insurers cannot collect information about certain risk factors at the proposal stage. Although this is usually intended to prevent discriminatory pricing, it also means that the insurer lacks information about

the composition of the risk pool which might be necessary or useful for its actuarial management. This point appears usually to be overlooked, and may be an unintended consequence of the drive against discrimination. Even in the longer term, insurers are likely to be specifically kept away from research data collected in future. In the U.K., the Biobank project is being set up, with large-scale funding, to recruit a cohort of 500,000 people aged 45–69, to obtain blood samples from all of them, and to study the genetic and environmental influences as they develop common disorders such as heart disease (www.biobank.ac.uk). The protocol specifically states: "Data from the project will not be accessible to the insurance industry or any other similar body".

What, then, are the general features of genetic epidemiology, if that must be the ultimate source of all the actuary's information? We must accept that epidemiological work is aimed largely at medical questions, and so the methods of medical statistics will figure largely. These may have a basis in common with actuarial science, for example in survival analysis, but there are differences of practice and emphasis. Actuarial models, even the ordinary life table, are in fact exceptionally demanding of the data, by requiring detailed, age-dependent risk estimates. Not many medical studies, especially into rare disorders in respect of which samples may be small, meet these demands.

General features of genetics

First, we must outline a few general features of genetics and genetic disease. We skip the details, which can be found in actuarial reviews such as Daykin *et al.* (2003), Doble (2001) or Macdonald (2003b), or much more fully in genetics texts such as Pasternak (1999) or Strachan & Read (1999).

(a) A gene is a region of DNA. Most genes encode the production of a particular protein, and, when prompted to do so by molecules that function as on/off signals, the gene product is produced. The extent to which genes are *expressed* in this fashion in individual cells depends on the tissue type and the environment that causes the signalling molecules to be produced or suppressed. These latter control networks are themselves mediated by other genes, often in immensely complex ways.

(b) A few, rare, disorders are associated with defects in a single gene. The great majority

are associated with complexes of variations in the many genes that participate in one of the biological pathways described above, also influenced by the environment. These common 'disorders' are called multifactorial. Roughly speaking, we currently know a fair amount about the increased risks associated with single-gene disorders, and very little about multifactorial disorders. The latter are the main target of large-scale prospective studies such as Biobank in the U.K..

(c) The burden of disease associated with a particular genotype results from:
   (1) the frequency with which it appears in the population;
   (2) the *age-related penetrance* of that genotype, meaning the probability that the disease will have appeared by any age x, all other decrements being excluded; and
   (3) the extent to which the symptoms may be treatable or avoidable.

   Not surprisingly, many genetic studies aim to estimate genotype frequencies and penetrances. And, although these are clearly the key components in any actuarial model of the disease, the question of whether effective treatments exist or not cannot be ignored in actuarial studies.

(d) Everyone has two copies of each gene (barring those on the X and Y sex chromosomes, which we will ignore here). 'Copies' need not mean identical stretches of DNA, but may mean slightly different versions of the gene, called *alleles*, some of which may be rare mutations and others of which may be common in the population. It is these variations that account for the variations between all people. One copy (allele) of each gene is inherited from each parent, which has consequences for the inheritance of diseases:
   (1) a single-gene disorder may cause a clear pattern of inheritance, either dominant (where one disease-causing allele over-rides any normal allele inherited from the other parent) or recessive (where a normal allele over-rides a disease-causing inherited from the other parents); or
   (2) a multifactorial disorder may show a very vague pattern of inheritance, hard to distinguish from shared familial environment, as the various disease-causing alleles at several genetic loci segregate independently from parents to offspring.

Genetic epidemiology

Genetic epidemiology is simple in principle. Just isolate a random sample of the population at risk of a particular disorder, and carry out a survival analysis of the experience. This is the idea behind the U.K. Biobank, but the fact that that needs a sample of 500,000 lives, and may still be limited in its power to detect moderate risk factors, hints at the difficulties. Sampling the whole population in order to capture the small proportion at risk of a rare disease is theoretically perfect, but wasteful of resources, so prospective studies are rare. Most are retrospective, often case-control studies. A feature of the latter is that they do not yield estimates of *relative risks*, which are exactly what actuaries would recognise as 'percentage extra morbidity', but only of odds ratios. Moreover, age is often regarded as a categorical variable, by dividing it into broad (often very broad) ranges, so that almost all age-related information is lost. (There seems to be no good reason why investigators could not ask for a date or year of birth, and themselves convert these to age ranges, but if this is actually done it is not usually obvious in the published studies.) Studies that do retain age-related information usually publish Kaplan-Meier estimates of survival probabilities, but often in the form of small graphs. For example, one work (Gutiérrez & Macdonald, in progress) looking at APKD found three sizeable studies which gave Kaplan-Meier estimates for the incidence of kidney failure associated with mutations in the APKD1 and APKD2 genes. In one case, the authors provided the actual numbers (ages and probabilities) shown in the graph; in the other two cases, the printed graphs remained the entire source of data. Unscientific as it may sound, this is evidently a common problem, as the following quotes show.

> "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." (Tudur *et al.*, 2001).

> "One rarely sees K-M tables reported because of their length. However, K-M curves are frequently seen. ... The cross hairs were added to the graphs for accurate straight-edge alignment." (Wesley, 2002).

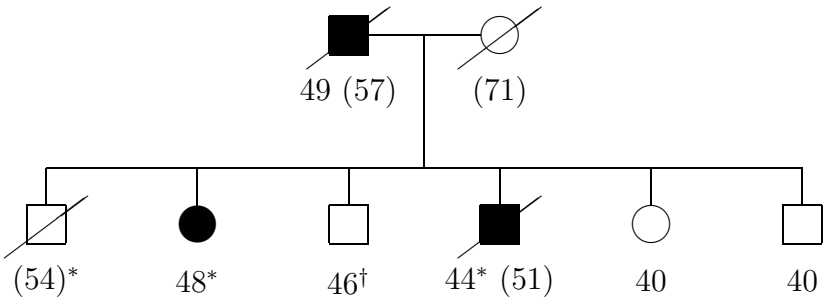


Figure 1: A hypothetical example of a pedigree. Squares are males, circles are females, and a slash denotes death. Affected individuals are shown as filled squares/circles. The age at onset or oldest observed age free of AD is shown, and age at death is given in brackets. An asterisk means that a person 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 (2002a).

Restricting ourselves to those genetic studies that give reasonable age-related information, what problems remain? We list some of the major ones below, see Macdonald (2003b) for more details.

(a) *Ascertainment bias* may be called the curse of genetic epidemiology. It arises because information about inherited risk must often be inferred from pedigrees; Figure 1 shows a hypothetical pedigree. The question arises: why has a particular family come to the attention of researchers? Much of genetic epidemiology depends on the idea of a *proband*, namely an affected person through whom the family was ascertained. The scope for bias is obvious, since a family which does carry the mutation, but in which by chance nobody develops the disease, will never be ascertained. This can be avoided by conditioning likelihoods on the probability of ascertainment through the actual probands(s), and there is an extensive though by no means finished literature on this subject (see Thompson (1993) or Hodge (2002)).

In studies of rare genetic disorders, the proband model may be inappropriate. Families may be chosen for study precisely because they have a large number of affected members, and it is then unclear to what extent the resulting risk estimates can be applied to other groups. This was true, for example, of the Breast Cancer Linkage Consortium (BCLC), a large number of families affected by hereditary breast cancer (BC), from which estimates of the risk associated with mutations in the BRCA1 and

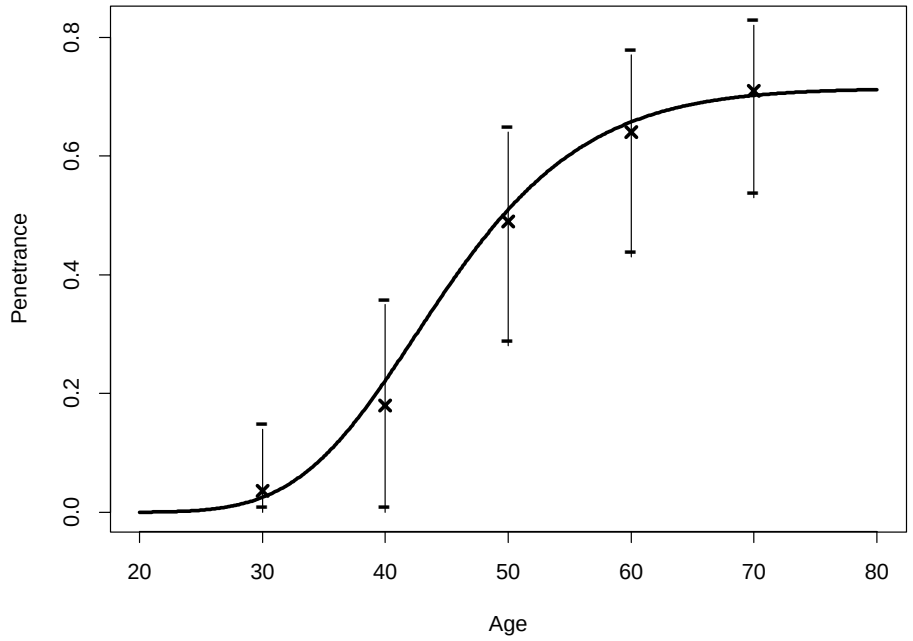


Figure 2: Observed values (×) and 95% confidence intervals of breast cancer penetrance associated with BRCA1 mutations, based on Ford *et al.* (1998). Also shown is the fitted function from Macdonald, Waters & Wekwete (2003a).

BRCA2 were obtained (Ford *et al.*, 1998) (this is the subject of the detailed case study in Section 5). This sample suggested penetrance of 80% at high ages (breast cancer, BRCA1 mutations), but later population based case-control studies suggested that 60% or 40% might be more accurate (see Lemaire *et al.* (2000), Subramanian *et al.* (2000), Macdonald, Waters & Wekwete (2003a)). Figure 2 gives an example of the kind of information available from such studies; note the small number of point estimates and the very wide confidence intervals.

If a woman who has had a genetic test for BRCA1 or BRCA2 applies for insurance in the U.K., she is presumably a member of a high-risk family, or else she would not have been tested, so should the higher, biased risk estimates be used? Would this be true in the U.S.A., where over-the-counter testing may be more widely available? More intriguingly, might the lower risk estimates based on population-based studies be universally applicable, even to the *unaffected* members of the very high-risk families, if those families were found by trawling the entire world?

(b) While some people in a family may be willing to have presymptomatic genetic tests,

often many are not, so the probability that they carry a mutation must be estimated using Mendel's laws, applied to those of their forbears known to carry mutations. Generally, the more unpleasant and untreatable a disease is, the lower is the take-up rate of genetic testing (Meiser & Dunn, 2000). For some diseases, carrier status may be inferred from positive disease status (for example, HD) but for other diseases which have rare familial forms (like BC) this is not so. Also, penetrance is rarely 100%, so absence of disease even at quite high ages usually does not imply mutation status either.

(c) Genetic disorders are very heterogeneous. Multifactorial disorders are so by their nature, but molecular analysis is revealing that single-gene disorders are too. Some disorders may be caused by mutations in one of several genes, possibly with quite different penetrances. Thus APKD is now known to have a severe form (caused by mutations in the APKD1 gene) and a less severe form (caused by mutations in the APKD2 gene); HD is known to vary in penetrance depending on the number of repeats of the CAG trinucleotide in the huntingtin gene; early-onset Alzheimer's disease (APKD) is known to be caused by mutations in one of three genes and so on (see Gui (2003), Gui & Macdonald (2002a), Gutiérrez & Macdonald (2002a, 2003)). Some genes, including BRCA1, have hundreds of known mutations, and research is gradually identifying different risks associated with mutations in different regions of the same gene. The actuarial problem is how this rapid shift from homogeneity towards heterogeneity might affect the assessment of reliability and accuracy, in the eyes of bodies like GAIC and the HGCA?

This brief survey can only give a small glimpse into the welter of information that is accruing, thanks to molecular genetics. How much of this will justify funding expensive epidemiological studies remains to be seen, and it is possible that in many instances we may end up with strong evidence of a genetic influence upon a disease, but no good quantitative knowledge of age-dependent risks.

4. Markov models

To a first approximation, dominant single-gene disorders divide the population into

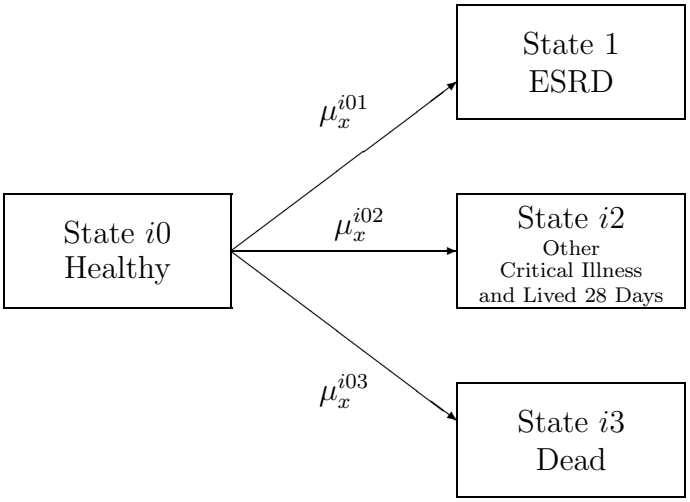


Figure 3: A model for APKD and Critical Illness insurance, in the i^{th} of several sub-populations representing genotype. Source: Gutiérrez & Macdonald (2003).

two distinct genotypes; mutation carriers and non-carriers. If the disorder has no cause except the gene mutation, non-carriers are at no risk at all. The discrete nature of the resulting risk groups is well-suited to the use of multiple-state models. As an example, consider APKD and critical illness (CI) insurance. Figure 3 shows a model in which a healthy person can progress to end-stage renal disease (ESRD, meaning kidney failure) because of APKD, which will result in a CI claim; or can claim for any other reason, including ESRD not resulting from APKD; or can die. There are two sub-populations (mutation non-carriers and carriers) labelled $i = 1$ and $i = 2$ respectively, and the difference between them is in the intensities of onset of ESRD. $\mu_x^{101} = 0$ because APKD is entirely genetic, while μ_x^{201} has to be estimated from epidemiological studies.

As mentioned in Section 3, some disorders, including APKD, may be caused by mutations in any of several genes. This is easily dealt with by defining separate sub-populations for carriers of mutations in each gene, usually ignoring multiple mutations because of their rarity. This could, in principle, be extended to allow for different mutations within the same gene (see Gutiérrez & Macdonald (2002a) for an example based on HD) but this will generally have to await much more detailed epidemiology. We note that most epidemiological studies to date assume homogeneity.

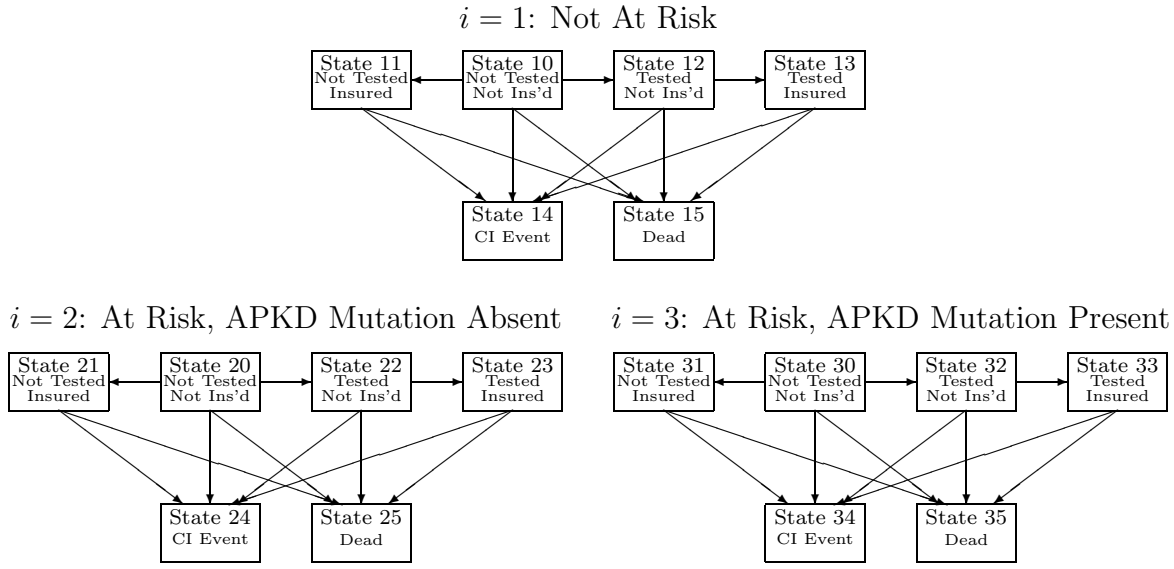


Figure 4: A Markov model of Critical Illness insurance allowing for family history of APKD and genetic testing. Source: Gutiérrez & Macdonald (2003).

To estimate the costs of adverse selection, we must extend the models to the purchase of insurance. An example of the simplest such model is shown in Figure 4. There are three sub-populations:

(a) persons with no family history, not at risk of APKD ($i = 1$);

(b) persons at risk of APKD because they have a family history, but who do not in fact have a mutation ($i = 2$); and

(c) persons at risk of APKD because they have a family history, and who do have a mutation ($i = 3$).

The members of the i^{th} sub-population start in state $i0$ in which they have not had a genetic test, nor have they bought insurance. From there they can simply buy insurance (move to state 01) or have a genetic test (move to state $i2$) and then perhaps buy insurance (state $i3$). At any time they can die or suffer a CI event. This model captures all the features we need:

(a) The size of the insurance market is determined by the rate at which insurance is bought, which can depend on age if that information is available.

(b) The incidence of genetic testing (screening in the whole population, or testing of at-

risk persons only) is represented by the rate of genetic testing in each sub-population.

(c) The mutation frequency is represented by the proportions in each sub-population. For example, APKD mutations occur in about 1 per 1,000 of the population, so at younger ages 0.1% would be at-risk carriers, 0.1% at-risk non-carriers and 99.8% not at risk. At older ages the mutation frequencies among healthy persons can be found by solving the Kolmogorov forward equations for the occupancy probabilities.

(d) The behaviour of 'adverse selectors' — both the probability that they buy insurance and the amount they buy — is represented.

(e) Each state in the model can be assigned to the appropriate underwriting class, depending on what information the insurer is allowed to use, and appropriate premiums can be calculated within each class.

(f) The model can easily be extended to allow for mutations in different genes, or more events such as lapsing insurance or buying more insurance (Pritchard (1997) and Subramanian *et al.* (2000) for example).

(g) Variations of the model can handle life, disability or long-term care insurance, or annuities (Tan (1997) for example). CI insurance is the easiest to model, since we need only rates of onset of the disorder. To model life insurance we need survival rates after onset, which often depend on duration as well as on age. There is little or no useful genetic epidemiology relating to disability.

The obvious difficulty is in choosing the intensities, especially those relating to behaviour. Research on attitudes to risk would be very helpful here.

The basic tools for handling multiple-state models are differential equations: Kolmogorov's forward equations for occupancy probabilities, and Thiele's equations for prospective reserves (Hoem, 1988). By solving Thiele's equations with suitable choices of premium rates and benefits, the expected losses with and without adverse selection are found, and the excess costs arising because of adverse selection are translated into an increased rate of premium. We give examples of these calculations in Section 5.

Multiple-state models give, at best, a discrete approximation to multifactorial disorders. However, it may in any case be some time before there are any useable risk estimates,

from studies like the U.K. Biobank. To date actuarial models have furnished only broad conclusions based on 'top-down' models of multifactorial disorders as an entire class. By 'top-down' we mean that extremely adverse assumptions are made about the genetic risk the incidence of genetic testing, and the extent of adverse selection. It may then be the case that the cost of adverse selection is small. Macdonald (1997, 1999) is an example in which the entire class of multifactorial disorders was considered, and Macdonald (2003a) treated the entire class of single-gene disorders similarly.

With a 'top-down' model is that we do not need detailed epidemiology of individual disorders, just enough information to be sure that our assumptions are extreme, but they are limited to seeking 'null results', in which extreme assumptions have small consequences. If that is not the case, we need a 'bottom-up' approach, in which we model each genetic disorder and aggregate the individual costs. This is more, but in the long run it is the only convincing approach.

The first 'bottom-up' models were those of Lemaire *et al.* (2000) and Subramanian *et al.* (2000), dealing with breast/ovarian cancer and life insurance, and Smith (1998), dealing with Huntington's disease (HD) and life insurance. Pokorski & Ohlmer (2000) also addressed breast cancer genetics in an insurance context. Others are Macdonald & Pritchard (2000, 2001) and Warren *et al.* (1999) on Alzheimer's disease and long-term care, Macdonald, Waters & Wekwete (2003a, 2003b) on breast/ovarian cancer and CI insurance, Gutiérrez & Macdonald (2003) on APKD, Gutiérrez & Macdonald (2002a, 2002b) on HD, and Gui (2003) and Gui & Macdonald (2002b) on early-onset Alzheimer's disease.

In the next section, we give a detailed example of the kind of study into adverse selection that is possible with fairly simple Markov models. This case study is based on Lemaire *et al.* (2000) and Subramanian *et al.* (2000).

5. An Example: the BRCA Mutation

Several years ago, two gene mutations that affect the likelihood of developing breast (BC) and ovarian (OC) cancer were discovered. Commercial tests to detect the presence of these mutations are now available. Women who learn through genetic tests that they are at higher risk of death for BC or OC cancer may purchase more life insurance, which to them looks inexpensive since it is priced at rates set for average risks. Women who learn they are at lower risk after a negative test may purchase less life insurance. These two forces combine to increase the aggregate mortality of the insurance purchasers. If insurers do not have access to the test results, they are unable to identify which women are at higher risk and which are not. They have to increase premiums for everyone, driving those at lower risk out of the pool. This creates a spiral of increasing prices and decreasing number of policies issued, which may threaten the financial solvency of the insurer.

Increased forces of mortality

The vast majority of BC and OC is the result of diet, lifestyle, environmental exposures, social interactions, and other factors, known and unknown. For instance, a late age at first childbirth and an early first menstruation slightly increase the likelihood of developing BC. Women with more pregnancies, or longer use of oral contraceptives, or who underwent tubal ligation or hysterectomy, have a reduced probability to develop OC. However, about six percent of breast and ovarian cancers are inherited. A small percentage of women (estimates range from one woman out of 833 to 2.3%) has a mutated dominant gene called BRCA1 or BRCA2. Women with a BRCA mutation are at extreme risk to develop BC or OC. Estimates of the probability to develop either of these cancers by age 70 are as high as 0.945, although this estimate was obtained from a selected group of women. The presence of a BRCA mutation in a 30-year old woman may reduce her life expectancy by 9.3 years.

Approximately one in nine women in the United States will develop BC in her lifetime; one in forty will die from the disease. Probabilities to develop BC, as a function of age and family history, have been obtained by medical researchers. For instance, table 1 indicates the predicted cumulative probability of BC for a woman who has a mother or sister affected, by age of onset of this first degree relative (FDR.) Onset is defined as the moment BC is diagnosed. Survival probabilities exhibit exponential decay: the annual probability that a woman affected with BC will die from the disease is 0.036, irrespective of the time since diagnosis and age at onset.

Table 1: Cumulative Probability of BC for a Woman who has One FDR Affected with BC, by Age of Onset of the Affected Relative

Age of Woman	Age of Onset of Affected Relative					
	20-29	30-39	40-49	50-59	60-69	70-79
29	0.007	0.005	0.003	0.002	0.002	0.001
39	0.025	0.017	0.012	0.008	0.006	0.005
49	0.062	0.044	0.032	0.023	0.018	0.015
59	0.116	0.086	0.064	0.049	0.040	0.035
69	0.171	0.130	0.101	0.082	0.070	0.062
79	0.211	0.165	0.132	0.110	0.096	0.088

OC is less prevalent, but deadlier: 1.8% of women will get the disease. The risk is multiplied by 5.4 in the presence of family history. Survival rates are low, but improving. In 1973, only 59.9% of the women who developed OC survived the first year after diagnosis. The five-year survival rate was 36%, the 20-year rate was 30.1%. In 1992, 78.3% of affected women survived their first year with OC. Applying Taylor's method to OC survival rates, the present-day five-year survival rate was estimated to be 50.9%, and the 20-year rate was 36.3%.

Estimates of the penetrance (the percentage of those with the gene mutation who will develop BC) of BRCA vary from 56% to 85%, depending on the selection bias in the population under study. This wide range of estimates is typical of the medical cancer literature. We wish our estimates of increased forces of mortality to be conservative from the insurer's perspective, i.e. our assumptions will tend to somewhat overstate the additional costs of life insurance. An average penetrance of 65% was selected as conservative. BRCA mutations not only increase the probability of developing BC, they also lead to earlier cancers: the age at onset of BC for women without the mutation is normally distributed around a mean of 69 years with a standard deviation of 15.4. With a BRCA mutation, the mean age at onset drops to 55.4, and the standard deviation is unaffected. Estimates of the likelihood to develop OC for a woman with a BRCA mutation vary widely, from 11% to 84%, depending on the type of mutation, the specific allele of BRCA1, and the population under study. An average of 40% seems conservative.

Based on these medical estimates, a double-decrement model was built to evaluate the increased force of mortality of a woman with a family history of BC or OC, or with a BRCA mutation. First, the survival probabilities for females given by the US Decennial Life Tables for 1989-91 were fitted to a Makeham distribution. Then, excess forces of mortality were calculated and fitted with a quadratic function. Table 2 presents the μ-ratio, the ratio of the force of mortality with family history or a gene mutation to the baseline force of mortality, for a 30-year-old woman, cancer-free at age 30. A second degree relative (SDR) is a grandmother or an aunt.

Table 2: μ-Ratios with a Family History of BC or OC, or with a BRCA Mutation, 30-year-old Woman, Age at Onset for BC: 20-29

Age	1 FDR-BC	1 SDR-BC	2 FDR-BC	1 FDR-OC	BRCA
31	1.0000	1.0000	1.0000	1.0302	1.0298
33	1.0345	1.0161	1.1051	1.1946	1.3543
35	1.0999	1.0465	1.3034	1.4011	1.8615
37	1.1822	1.0848	1.5518	1.5958	2.4323
39	1.2627	1.1225	1.7927	1.7350	2.9322
41	1.3385	1.1580	2.0159	1.7070	3.2351
43	1.3004	1.1391	1.9045	1.5812	2.9300
45	1.2976	1.1358	1.8999	1.6926	3.0133
47	1.3026	1.1362	1.9167	1.8143	3.1367
49	1.3174	1.1414	1.9586	1.9083	3.2691

Excess mortality can reach 100% in some cases of family history of BC, and 225% for a woman with a BRCA mutation. It seems to be common practice among insurers to accept at ordinary rates applicants with a force of mortality up to 150% of aggregate. Consequently, while some females with a family history of cancer can be accepted at standard rates, others need to be quoted sub-standard rates. Depending on the underwriting policy of the

company, females with a gene mutation can possibly be covered, at a rate incorporating a severe mortality surcharge. Note that the common assumption that a given disease simply multiplies forces of mortality by a constant (*constant frailty hypothesis)* does not apply in the case of BC and OC. Table 3 shows that these mortality increases are comparable to increases resulting from common risk factors.

Table 3: Mortality Ratios for Common Risk Factors

Risk Factor	Measurement	μ-ratio
High systolic blood pressure	158-167 (men)	2.06
High systolic blood pressure	178-187 (women)	2.78
Diabetes mellitus	Men	2.50
Build	40 percent overweight (women)	1.62
Build	60 percent overweight (men)	2.60
Epilepsy	All types	2.78
Alcoholism	5 drinks a day	3.00
Smoking	Average (men)	1.70
Smoking	40 cigarettes/day (men)	2.00
HIV	35-year-old male	50.00

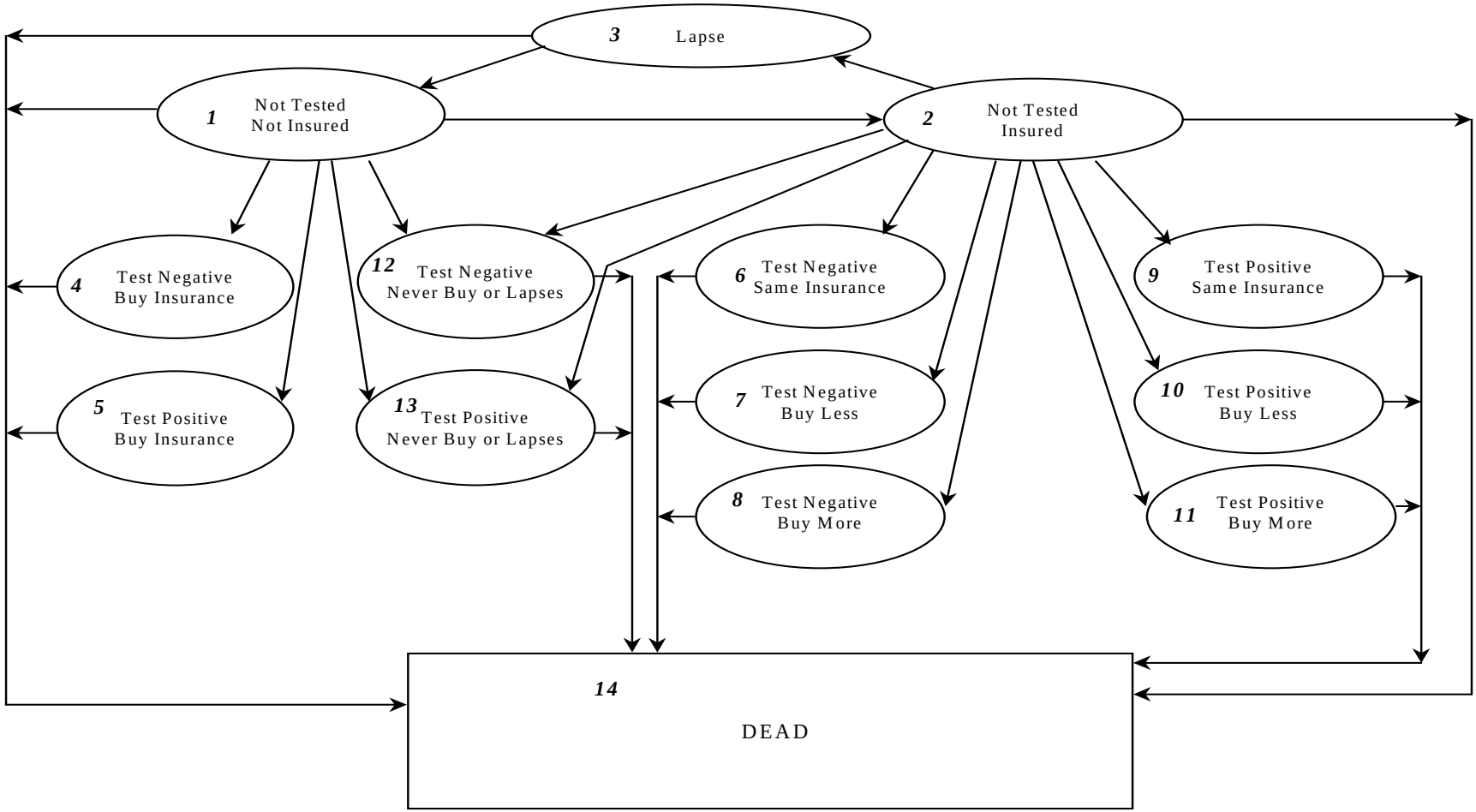
A Markov Model

A continuous-time, discrete-state Markov model is developed here to represent the actuarial environment of genetic screening. The model, shown in figure 1, decomposes the history of an individual into a series of discrete states (represented by ellipses), which analyze single premium term insurance purchasing and genetic testing decisions. At all times, every individual is assigned to one and only one state. Transitions (represented by arrows) from one state to another can occur at any time. Forces of transition only depend on the state currently occupied and not on past history. At time 0, a woman may be in either state 1 or state 2. A woman in state 1 has not been tested for BRCA mutations and has no insurance. A woman in state 2 has not been tested, but has insurance. From state 1, six transitions are possible: a woman can remain untested and purchase insurance (state 2); she can test negative and buy insurance (state 4) or remain uninsured (state 12); she can test positive and buy insurance (state 5) or remain uninsured (state 13); she can die before getting tested or becoming insured (state 14). She can also remain in state 1. Correspondingly, ten future states are possible from state 2, including the possibility of lapsing the policy. Most transitions have cash flow implications. Transitions into states 2 and 4 through 11 imply insurance premium payments. Transitions from states 2 and 4 through 11 into state 14 imply insurance benefit lump sum payments. We assume that policies are purchased through net single premiums, and that the demand for insurance is inelastic to price. Only the applicant has full information about her state.

One such Markov model can be defined for every age of the population at time zero, and for every possible family history. We consider three initial ages (30, 40, and 50) and four family histories (no BC or OC in the family; one FDR with OC; one FDR with BC, onset age 20 to 29; two FDR with BC, onset ages 20 to 29 for both), resulting in 12 subgroups. At time zero, women are assumed to be unaffected by BC or OC.

Figure 5

Markov Model: Heritable Breast Cancer and Adverse Selection



Thiele's Equations

The (continuous time) force of transition at time *t* from state *j* to state *k*, for subgroup *i*, is denoted $\mu_t^{i,j,k}$ (in the sequel superscript *i* will be omitted.) For each state, we wish to calculate, under a variety of assumptions, the actuarial present value of future insurance benefits, incorporating mortality and interest. This benefit reserve is a liability to the company. As the reserves for the various states are dependent, their values can only be found by solving a set of differential equations that generalizes Thiele's equation for benefit reserves. One differential equation can be written for each state for which there is an outward transition. The equation for state *j* is written

$$\frac{d}{dt}V_t^{(1)j} = \delta_t V_t^{(1)j} - \sum_{k \neq j}(b_t^{jk} + V_t^{(1)k} - V_t^{(1)j})\mu_t^{jk} \qquad (1)$$

where $V_t^{(1)j}$ = benefit reserve for state *j* at time *t*, δ_t = force of interest at time *t*, and b_t^{jk} = payment due upon transition from state *j* to state *k*.

Usually, Thiele's equation includes a positive premium rate term. This term is not present in equation (1) since policies are purchased with net single premiums. The interpretation of the differential equation is as follows: at all times, the reserve increases through interest accrual. Upon a transition from *j* to *k*, a benefit b_t^{jk} might be paid. This would happen only for a transition into State 14, when a death benefit is to be paid to a beneficiary. Switching from state *j* to state *k* also implies the release of the reserve for state *j* and acquiring the reserve for state *k*. The amount between parentheses is the net sum at risk.

This set of differential equations can be solved backward recursively, using the boundary conditions $V_{t^*}^{(1)j} = 0$, where *t** is the ending time for the period under consideration. At time *t**, the company no longer needs to hold funds aside for this policy, because the policy term and the corresponding financial obligation has ended. We solve the set of differential equations using a mathematical programming package.

We are interested in the *cost of adverse selection* attributable to the availability of testing, under conditions where women have access to genetic test results, but insurers either have no access or are prohibited from using that information in underwriting. To calculate this, we first solve the differential equations, assuming no allowed use of genetic testing by insurers. Women flow through the system at the baseline transition rates and experience mortality at rates based only on their family history, the information that the insurer uses for pricing purposes. This first solution gives us the expected present value of benefits in the "no use of genetic testing" case. We then solve the same equations, assuming genetic testing use is allowed. Women flow through states 1 to 13 at the same baseline transition rates. Net single premiums are paid in the same way. However, the transition rates into state 14 will now be mortality rates corresponding to the woman's BRCA status, if she is tested (baseline mortality if negative, BRCA mortality if positive,) and to her family history, if she remains untested. This solution provides the expected present value of benefits in the full information case. The ratio of the two measures yields the cost multiplier of adverse selection, the ratio of what the true risk is to what is claimed and charged.

$$Cost\ Multiplier\ of\ Adverse\ Selection = \frac{EV(insurers\ use\ full\ information)}{EV(insurers\ use\ allowable\ information)}$$

Benchmark Assumptions

Estimates are necessary for each force of transition. Transitions into state 14 reflect mortality rates that differ by age, family history and BRCA status. The other transitions involve a combination of testing behavior, test results, and insurance purchasing behavior. The following forces were selected for our baseline calculations.

- *Behavior before testing*. From industry figures, we estimate the rate of insurance purchase $\mu_x^{1,2}$ at 5%, the lapse rate $\mu_x^{2,3}$ at 5%, and the rate of re-entry $\mu_x^{3,1}$ into state 1 at 25%.
- *Rate of genetic testing r*. Very few women get tested presently: only 500 women have been tested at the University of Pennsylvania since the test became available, late in 1996. The test is very expensive ($2,400) and the cost is not expected to decrease dramatically, as one laboratory owns the patent. The testing rate may depend on family history; however a uniform rate enables us to compare adverse selection costs across family histories. Therefore, we select a testing rate of 5%. This rate may be considered too high for women with no family history of BC or OC and too low for women with 2 FDR with BC, but for a population of women, this rate is conservative.
- *Force of interest*. We assume a constant force of interest of 5%.
- *Test results*. The probability *p* that a test result will be positive depends on individual characteristics: a woman with two FDR affected by BC is much more likely to have the BRCA gene mutation than a woman with no family history of the disease. The value of *p* is found by introducing the constraint that expected benefits need to be equal in two cases: (1) no genetic testing; and (2) women get tested but their insurance purchase decisions are not affected. These constraints yield the following probabilities of a positive test result:

  p=0.005 no family history

  p=0.08 one FDR affected with OC, age at onset unknown

  p=0.15 one FDR affected with BC, onset age 20-29

  p=0.40 two FDR affected with BC, both with onset age 20-29
- *Changes in insurance benefits.* We assume the benchmark amount of term insurance to be $1. A woman buying "less insurance" always reduces her benefit amount from $1 to $0.50; a woman buying "more insurance," increases her benefit amount to $2, $4, or $10, varied in a sensitivity analysis.
- *Insurance purchase probabilities.* Insurance decisions are assumed to occur shortly after the test result is provided. Benchmark probabilities were selected as follows, assuming a high degree of inertia among tested women:
  - If uninsured and test positive: P(buy insurance) = 0.25
    P(not buy) = 0.75
  - If uninsured and test negative: P(buy insurance) = 0.05

P(not buy) = 0.95

° If insured and test positive: P(more insurance) = 0.27
P(same insurance) = 0.70
P(less insurance) = 0.02
P(lapse policy) = 0.01

° If insured and test negative: P(more insurance) = 0.02
P(same insurance) = 0.75
P(less insurance) = 0.18
P(lapse policy) = 0.05

Transition rates are then obtained by multiplying the appropriate rates and probabilities. For instance, $\mu_x^{1,12} = 0.95r(1-p)$..

Cost of Adverse Selection, by Family History

The model was run, using the baseline behavioral assumptions, for the four family histories under consideration. Table 4 shows adverse selection costs for a woman with no family history. Adverse selection invokes only a small cost in this case because the probability of having the mutation for a woman with no family history of BC or OC is only 0.005. Insurance companies should not be concerned with restrictions of the use of genetic testing information for women with no family history.

Table 4: Costs of Adverse Selection for a Woman with No Family History of BC or OC, Insured at Time 0

		Term			
Age	Increased Benefit	5	10	15	20
	$2	1.0006	1.0021	1.0039	1.0054
30	4	1.0009	1.0030	1.0056	1.0077
	10	1.0016	1.0057	1.0106	1.0145
	$2	1.0008	1.0026	1.0042	1.0046
40	4	1.0012	1.0037	1.0060	1.0066
	10	1.0023	1.0070	1.0112	1.0123
	$2	1.0004	1.0013	1.0020	1.0021
50	4	1.0006	1.0019	1.0029	1.0030
	10	1.0012	1.0035	1.0054	1.0056

Table 5 presents adverse selection costs for a woman with one FDR with BC, age at onset 20 to 29. Costs rise as a function of the selected increased benefit of women who get tested and exceed 10% in some cases. The most expensive aspect of adverse selection results from the women who select high benefit levels following a positive test. The lapsing behavior of the women who test negative has much less of an impact. The costliest part

of adverse selection is higher insured benefits. This result provides some support to regulations restricting the benefit amounts that can be obtained without having to disclose genetic test results.

Table 5: Costs of Adverse Selection for a Woman with One FDR Affected with BC, Age at Onset: 20-29, Insured at Time 0

		Term			
Age	Increased Benefit	5	10	15	20
	$2	1.0065	1.0177	1.0294	1.0386
30	4	1.0126	1.0369	1.0626	1.0825
	10	1.0300	1.0893	1.1495	1.1935
	$2	1.0144	1.0380	1.0544	1.0523
40	4	1.0233	1.0632	1.0925	1.0928
	10	1.0489	1.1322	1.1924	1.1949
	$2	1.0079	1.0197	1.0271	1.0250
50	4	1.0128	1.0331	1.0468	1.0449
	10	1.0269	1.0698	1.0985	1.0952

For the most part, adverse selection costs increase with duration, because longer terms give women more opportunities to get tested and increase their insured benefits. Occasionally, for some 40- and 50-year old women, the costs for a 15-year period exceed the costs for a 20-year period. Recall that these are ratios of expected losses. As these women age, they become more vulnerable to other causes of mortality, thus increasing both numerator and denominator and somewhat decreasing overall adverse selection costs.

The relationship between age and the cost of adverse selection is not monotonic. Of the three initial ages, adverse selection costs are always higher for a woman age 40. Women who are cancer-free at the age of 30 are relatively unlikely to develop BC or OC before the age of 50, and even less likely to die from the disease during that 20-year period. Women who are 50 are more prone to develop cancer, but also more prone to die from other causes, so that the adverse selection cost, which is a ratio, is lower.

Table 6 shows adverse selection costs for a woman with two FDR with early BC onset. Adverse selection costs are not higher than in the previous case. These costs evaluate the amount of information is present in the BRCA test that is not present in family history; there is a substantial probability that the gene mutation is present in the family when family history is strong.

Table 6: Costs of Adverse Selection for a Woman with Two FDR Affected with BC, Ages at Onset: 20-29, Insured at Time 0

		Term			
Age	Increased Benefit	5	10	15	20
	$2	1.0089	1.0168	1.0237	1.0294
30	4	1.0186	1.0421	1.0631	1.0787
	10	1.0445	1.1034	1.1510	1.1819

	$2	1.0259	1.0578	1.0749	1.0667
40	4	1.0419	1.0956	1.1264	1.1183
	10	1.0848	1.1873	1.2413	1.2269
	$2	1.0143	1.0296	1.0365	1.0307
50	4	1.0234	1.0510	1.0652	1.0584
	10	1.0479	1.1030	1.1298	1.1171

Finally, we consider the financial consequences when insurers do not incorporate the correct family history information in the pricing process. This could result from either a ban on the use of family history in underwriting or from policyholder fraud. Our previous results assumed that women report their family history truthfully and that insurers are allowed to use family history in underwriting. Now assume a woman with 2 FDR with BC reports no family history of BC or OC, and the insurer either fails to detect this fraud or is not permitted to use the truthful information provided. Table 7 reveals huge adverse selection costs as compared to table 6, in which an insurer correctly underwrites a woman with this family history. It is of crucial importance for insurers to request detailed family information (including age at onset) during the underwriting process, and to investigate the applicant's statements vigorously.

Table 7: Costs of Adverse Selection for a Woman with Two FDR with BC, Ages at Onset: 20-29, Insured at Time 0, Priced at No Family History of BC or OC

		Term			
Age	Increased Benefit	5	10	15	20
	$2	1.2356	1.5149	1.7253	1.8490
30	4	1.2813	1.6364	1.9312	2.1263
	10	1.4170	1.9933	2.5300	2.9260
	$2	1.2536	1.5176	1.6803	1.6947
40	4	1.3076	1.6565	1.8963	1.9461
	10	1.4678	2.0639	2.5243	2.6703
	$2	1.1469	1.2909	1.3704	1.3669
50	4	1.1878	1.3865	1.5138	1.5339
	10	1.3090	1.6670	1.9305	2.0151

Conclusions

Should the insurance industry oppose any ban on the use of genetic testing results in underwriting, to avoid adverse selection? Should legislators enforce some degree of subsidization among policyholders, in order not to penalize people because of their bad luck in the genetic lottery, a situation for which they are not responsible? The answer to these questions may depend on the cost of adverse selection to insurance industry, and on the likelihood that it may threaten the solvency of some. We have attempted to provide some actuarial insight in the debate by quantifying the impact of adverse selection.

All our results have been obtained under conservative assumptions. Also our calculations assume that genetic testing leads to no medical benefits in the form of improved risk reduction. There is hope that women found to carry BRCA mutations can reduce their risk of BC mortality by increased surveillance through higher frequency of mammograms, prophylactic mastectomy, or chemoprevention with tamoxifen. The gain in life expectancy for a 30-year old woman with the BRCA mutation who choose to undergo prophylactic mastectomy and oophorectomy could be as high as 5.3 years. Therefore we believe that all figures are cautious upper bounds of adverse selection costs.

As Tables 4 to 6 illustrate, it is only in a few cases (20-year term, family history of BC with early age at onset, large benefit amounts) that the cost exceeds 10%. The problem resulting from very large benefit amounts could be alleviated if insurers were allowed to use genetic test results for the underwriting of such policies, in return for a ban on the use of tests for policies with a reasonable amount.

Under our approach, the average adverse selection cost in a portfolio is expected to be way below 10%. So, this cost is likely to be compensated by the overall long-term trend of decrease in mortality rates (a factor not introduced in our calculations) that stands around 0.64 percent per year these days. Therefore we believe that adverse selection is a problem that is controllable by insurers.

This conclusion only holds if companies apply very tight underwriting standards. In the application process, prospective insureds need to provide the detailed family history of all their first-degree relatives, with ages at onset of any cancer. Applicants' statements need to be carefully checked by underwriters. If companies fail to correctly identify the family history of the applicant, or if by law insurers cannot use family history as a substitute for genetic testing, table 7 shows that adverse selection costs could become unbearable.

Note that our conclusions are only valid for life insurance contracts. Life insurance underwriting is performed only once, renewal is automatically guaranteed without further evidence of insurability, and premiums are usually guaranteed for a long period. Health insurance would require another model, incorporating continual updates of the insured's health status and subsequent premium adjustments.

6. New models and questions

In this section we focus on some recent and continuing work into several of the disorders listed by the ABI as being of potential significance for insurance, namely:

(a) Huntington's disease (HD);
(b) early-onset Alzheimer's disease (EOAD); and
(c) multiple endocrine neoplasia type 2 (MEN2).

These disorders throw up a range of different challenges, both for the actuary and for those making and implementing policy, such as GAIC and the proposed HGCA. In addition, we will describe briefly a model for coronary heart disease (CHD) which is a first step in the modelling of complex multifactorial disorders.

Huntington's disease

HD is a degenerative brain disorder caused by an expansion of a certain trinucleotide repeat (CAG) in the huntingtin gene. A 'repeat' means that the DNA bases are repeated several times at a certain location in the gene — CAG-CAG-CAG-CAG. The number of repeats varies from person to person; normally it is 10–20, but once it reaches 35 or so, HD results. The repeat number is unstable, meaning that the reproduction of this region of the gene during the production of sperm and eggs is error-prone, and on balance it tends to expand. Important features of HD are:

(a) it is highly penetrant, so much so that it is often used as the canonical example of a highly penetrant late-onset disorder;
(b) it is almost always fatal, and no treatment is known;
(c) the penetrance is strongly correlated with the CAG repeat number; the more CAG repeats there are, the lower is the age-at-onset distribution; and
(d) the disease is progressive, passing through several stages that may be recognised qualitatively but for which no quantitative studies exist.

Smith (1998) proposed a model of HD, that was used as the basis of the ABI's application to GAIC in respect of life insurance (see Section 2). However this was based on early studies and did not allow for the association between CAG repeat number and

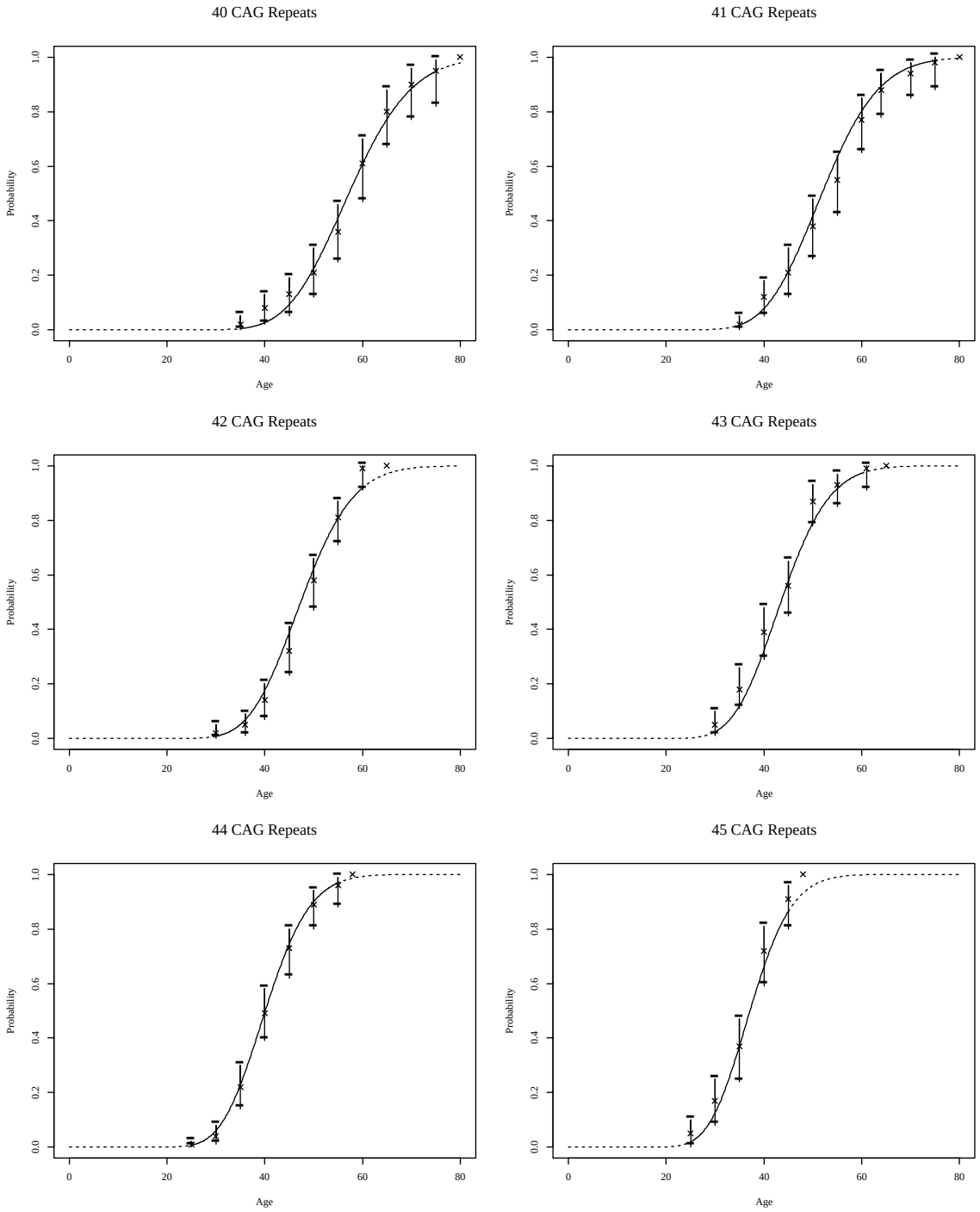


Figure 6: Penetrance estimates of onset of HD with 40–45 CAG repeats (crosses) and 95% confidence intervals, from Gutiérrez & Macdonald (2002a), based on data from Brinkman *et al.* (1997). Also shown are fitted penetrance curves.

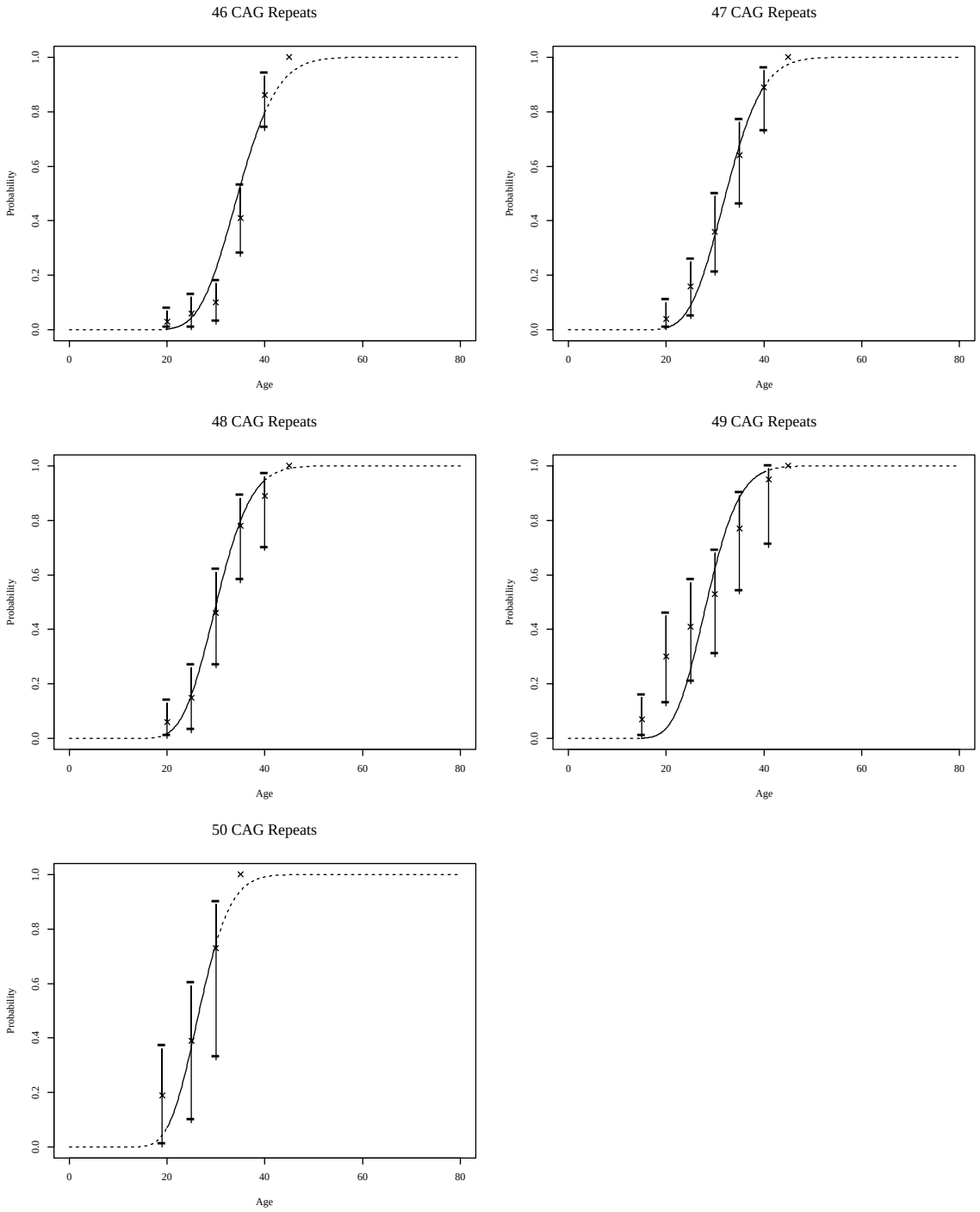


Figure 7: Penetrance estimates of onset of HD with 46–50 CAG repeats (crosses) and 95% confidence intervals, from Gutiérrez & Macdonald (2002a), based on data from Brinkman *et al.* (1997). Also shown are fitted penetrance curves.

penetrance. The main study of the latter is Brinkman *et al.* (1997). Figures 6 and 7 show Brinkman's point estimates of age-related penetrance for 40–50 CAG repeats, as well as the following fitted model from Gutiérrez & Macdonald (2002a):

$$\text{Penetrance at age } x = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} \exp(-t\theta) dt \tag{1}$$

where R is the CAG repeat number, $\alpha = 48.1685 - 0.376508R$, $\theta = 0.051744R - 1.49681$ and $x \geq 0$. Experiment suggests this model gives reasonable results when extrapolated down to 36 CAG repeats. Note, a most important point, that Brinkman *et al.* (1997) included a table with the actual numerical values of the estimates shown in the graphs. Basic as it seems, this is most unusual.

Using the fitted model, premiums for critical illness (CI) and life insurance can be found, and estimates of the cost of adverse selection, along the lines of Sections 4 and 5. Particular problems included the following (for details see Gutiérrez & Macdonald (2002a, 2002b):

(a) The stage at which a CI claim would be admitted is unclear, though it is certainly some time after onset. Based on qualitative descriptions of three main stages of the disease, given by Harper (1996), an accelerated lifetime model was applied to data on survival after onset, which were available.

(b) The cost of adverse selection depends on the distribution of CAG repeat numbers in the population at risk, namely new-born children in families at risk of HD. This was modelled by back-fitting from cross-sectional data on the distribution of CAG repeat numbers in adult populations.

Tables 8 and 9 show some examples of level net premiums for CI insurance, expressed as a percentage of a standard premium (basis given in Gutiérrez & Macdonald (2002b)). This assumes that a claim is paid when the second of the three stages of disease progression is reached, which is relatively onerous but might be the case under some policies. Table 8 shows 40–45 CAG repeats, while Table 9 shows 36–39 CAG repeats, based on the extrapolated model.

Table 8: Level net premium for level CI cover for persons with a known HD mutation, as a percentage of the premium for standard risks. Claims arising at Stage 2 of Harper's progression (Harper, 1996).

			Premium as Percentage of Standard										
Sex of	Age at	Policy	Number of CAG Repeats										
Applicant	Entry	Term	40	41	42	43	44	45	46	47	48	49	50
	(Years)	(Years)	%	%	%	%	%	%	%	%	%	%	%
Female	20	10	100	101	107	125	176	294	523	911	1,487	2,249	3,168
		20	111	141	221	389	682	1,118	1,690	2,373	3,125	3,907	4,680
		30	168	270	442	683	978	1,311	1,667	2,040	2,424	2,816	3,208
		40	246	361	502	658	828	1,013	1,215	1,437	1,676	1,928	2,185
	30	10	115	154	249	429	705	1,066	1,480	1,913	2,336	2,732	3,091
		20	183	308	515	800	1,137	1,492	1,837	2,154	2,434	2,676	2,882
		30	268	406	578	771	975	1,181	1,380	1,565	1,732	1,880	2,007
	40	10	180	274	405	559	718	871	1,011	1,136	1,246	1,341	1,424
		20	298	443	604	760	899	1,018	1,117	1,199	1,268	1,324	1,372
	50	10	247	330	414	494	567	630	685	733	774	809	840
Male	20	10	100	102	112	143	231	433	828	1,495	2,485	3,796	5,376
		20	115	159	275	519	942	1,574	2,403	3,392	4,482	5,614	6,736
		30	180	300	503	787	1,136	1,529	1,951	2,392	2,848	3,312	3,776
		40	242	356	495	651	821	1,007	1,211	1,435	1,677	1,931	2,190
	30	10	120	171	297	534	900	1,376	1,923	2,494	3,054	3,577	4,051
		20	192	328	556	869	1,239	1,629	2,009	2,358	2,666	2,932	3,159
		30	256	385	546	728	921	1,117	1,306	1,483	1,642	1,782	1,904
	40	10	179	273	403	556	714	866	1,005	1,129	1,238	1,333	1,415
		20	271	396	535	671	792	896	982	1,054	1,114	1,164	1,205
	50	10	216	281	348	411	468	518	562	599	632	660	684

Table 9: Level net premium for level CI cover for persons with a known 'intermediate allele' HD mutation (36–39 CAG repeats), as a percentage of the premium for standard risks. Claims arising at Stage 2 of Harper's progression (Harper, 1996).

		Premium as Percentage of Standard							
		Females				Males			
Age at	Policy	No. of CAG Repeats				No. of CAG Repeats			
Entry	Term	36	37	38	39	36	37	38	39
(Years)	(Years)	%	%	%	%	%	%	%	%
20	10	100	100	100	100	100	100	100	100
	20	100	100	100	102	100	100	100	103
	30	100	101	104	120	100	101	105	124
	40	101	105	121	165	101	104	121	163
30	10	100	100	100	103	100	100	101	104
	20	100	101	105	125	100	101	106	128
	30	101	105	124	174	101	105	122	169
40	10	100	101	106	127	100	101	106	127
	20	101	107	130	191	101	106	126	178
50	10	101	107	130	177	101	106	123	160

The most obvious feature is that while premiums are unfeasibly high for about 42–43 and more CAG repeats, except for shorter terms, cover should be available for quite reasonable extra premiums for people with fewer CAG repeats. (CI insurers typically decline risks greater than about 300% to 350% of the standard premium rate.)

The most interesting feature, however, relates to how a moratorium might be applied in practice. In the U.K., it is still permitted to use family medical history without restriction, so this gives the premium rate that should be paid by a person from an at-risk family, whose genetic test result (if any) is to be disregarded under the moratorium. The attitude of insurers, however, is that they will grant standard rates to someone who can show that they *do not* carry the mutation, as a result of a genetic test. Arguably this is in breach of the moratorium, but it seems to be unequivocally to the advantage of the applicant so this is not controversial.

Table 10 shows the premiums for a person who has a parent or sibling with HD (or known to carry the mutation) but who has not themselves had a genetic test. (This table also shows the effect of CI claims being admitted at the later third stage of the disease,

Table 10: Level net premiums for CI cover as a percentage of the premium for standard risks, for persons with a family history of HD (affected parent or sibling).

		Claims arising			
Age at	Policy	At Stage 2		At Stage 3	
Entry	Term	Females	Males	Females	Males
(Years)	(Years)	%	%	%	%
20	10	263	380	132	156
	20	503	684	246	311
	30	480	549	289	323
	40	388	387	268	266
30	10	266	320	137	148
	20	335	358	195	204
	30	296	284	203	197
40	10	172	171	115	115
	20	202	188	142	136
50	10	128	122	107	105

giving a considerable reduction.) So, for example, a female age 20 would pay 268% of standard rates for a 10-year policy. From Table 8, this exceeds the premium payable by a woman with 44 or fewer CAG repeats. Such a person could ask that their genetic test, adverse because it shows them to carry a mutation, *should* be taken into account because it would be to their advantage; the premium would be lower than that based on family history alone. If that were granted, the premium based on family history would increase substantially, because that underwriting group now includes only those people with 45 or more CAG repeats. That increase could make it advantageous for a woman with 45 CAG repeats to ask for a premium based on that fact, and so on.

It seems reasonable, on first sight, to use genetic test results when they are to the applicant's advantage. Indeed, insurers have used this to argue that genetic tests will allow non-carriers to obtain cheaper insurance. This might not be problematic for a completely homogeneous disorder, where one unique mutation leads to disease, but for a very heterogeneous disorder such as HD, the same logic could extend to the use of some adverse test results, as we have seen. In terms of any moratorium, this is probably both unexpected and unacceptable.

Early-onset Alzheimer's disease

In EOAD, the actuarial interest focusses on the difficulty of obtaining any estimates of onset rates at all. Three genes are implicated; Amyloid Precursor Protein (APP), Presenilin-1 (PSEN-1) and Presenilin-2 (PSEN-2). The estimation problem for PSEN-1 mutations has already been discussed at an ASTIN colloquium (Washington in 2001) and in ASTIN Bulletin (Gui & Macdonald, 2002a). It seems not to have been covered in the genetic epidemiology literature. Gui & Macdonald (2002b) applied the results to CI and life insurance, and since the methodology was similar to that used for HD we refer the reader there for details. One point of interest is that life insurance premiums could be higher for a carrier of a PSEN-1 mutation than for a carrier of a HD mutation, even though HD is usually thought of as the most severe single-gene disorder. The reason is that mortality after onset of EOAD is higher than after onset of HD, and in fact the latter is lighter than might be imagined.

The APP and PSEN-2 genes are more problematic. Mutations in PSEN-2 are very rare, and for some time were only known to exist in a single extended kindred. The ABI left it off its list, but did include PSEN-1 and APP. Gui & Macdonald (2003b) found insufficient data in published pedigrees to get acceptable estimates of onset rates for EOAD associated with either APP or PSEN-2 mutations and, as for PSEN-1, no other estimates are available in the literature.

The interesting question, therefore, is how the 'actuarial relevance' of APP mutations might be demonstrated, in terms of the GAIC or HGCA processes? Some possibilities are:

(a) Defer consideration until such time as the epidemiological rates can be obtained.

(b) Use very broad-brush comparisons between the effects of APP mutations and some other gene mutations, such as PSEN-1, for which more reliable evidence is available. This is not unlike the method of graduating sparse mortality experiences by reference to a standard table. It would presumably require a joint model of the penetrance of PSEN-1 and APP mutations.

(c) Rely on the opinion of medical experts.

This is a problem that will occur more often in future, if underwriting becomes more 'evidence-based'. This might mean being able to justify decisions either in advance of underwriting, or if afterwards, at least without forcing the applicant to bring a court case. Some of the ALRC's recommendations are moves in this direction. The relative value of qualitative and quantitative risk information will be tested. For example, it is arguably quite obvious that APP mutations are highly penetrant and lead to early death. Given almost any threshold, such as the +50% additional mortality as originally used by GAIC, it would clearly be 'actuarially relevant'. But if a model could not be calibrated, how then would insurers calculate a reasonable premium, that could withstand challenge from GAIC or the HGCA or a court?

It would be unfair and absurd to force insurers to pretend that no additional risk existed, when it was clearly very high but not easily quantified any more accurately than that. Very rough estimates may then be allowed for premium calculation. The *quid pro quo* must be that more accurate underwriting is required, wherever and whenever the evidence exists to support it, within the limits of the individual insurer's underwriting philosophy. Then a body such as GAIC or the HGCA might adjust its criteria in the light of the available evidence, asking for sharper numbers when good evidence exists and allowing more leeway otherwise. It will be interesting to see how the reformed GAIC approaches this problem.

Multiple endocrine neoplasia type 2

Gui & Macdonald (2003a) investigated MEN2, and rather like APP and PSEN-2 found that the literature simply did not support any actuarial model. However, it did lead to quite a different conclusion from that above.

MEN2 is a dominantly inherited disorder characterised by tumours in the various endocrine glands, of which the most significant is medullary thyroid carcinoma (MTC). It has several forms, of which we pick out two: MEN2A accounts for about 15% of MTC, with onset at ages 21-38; and MEN2B accounts for about 3% of MTC, with onset at ages 12–23. They are associated with certain mutations in the RET gene, for which a test exists.

Provided it is detected before onset of symptoms, MEN2 can be treated with high success by prophylactic removal of the thyroid gland. This operation is not drastic, and can normalise mortality. Thus MEN2 fulfils the criteria for a screening program, and it is recommended that screening should be carried out at ages 3–5. Genetic and biochemical tests are available, but genetic testing leads to much earlier detection and better outcomes. So why is MEN2 on the ABI's list of disorders of significance to insurance?

MEN2, like the other disorders on the list, has dominant inheritance, late onset and high penetrance, all the factors that might affect insurance. But, unlike HD, for which no treatment exists, or familial breast cancer, for which only drastic surgery is the only treatment on offer, MEN2 is eminently treatable. The greatest risk to people in at-risk families, therefore, is that they are not tested, and mutations go undiscovered.

One of the primary concerns of medical professionals is that worries about insurance might make people choose not to have beneficial tests or treatment, either for themselves or their children, and MEN2 is a case in point. It seems that the industry's main aim should be to ensure that its practices do not create obstacles to effective treatments. In the U.K., it seems that the ABI does not intend to pursue any application to be allowed to use genetic test results in respect of MEN2, at least for the time being.

Coronary heart disease

Coronary heart disease (CHD) is a major cause of serious illness and death in developed countries. Underwriters have long recognised that family history is significant, although whether that means genetic or shared environmental risk has been impossible to say. With a few rare exceptions, such as the LDL receptor gene, there is no clear Mendelian pattern of inheritance that would suggest a single gene at work; this is a multifactorial disorder.

CHD follows the narrowing of the arteries that supply the heart muscle, by the deposition of fatty plaques. This leads to symptoms such as angina, or if an artery becomes blocked, to myocardial infarction (MI, meaning a heart attack) when some of the heart muscle dies, and possibly the patient too. MI is the 'endpoint' that might trigger a CI or life insurance claim, but it is at the end of a long process. Many significant

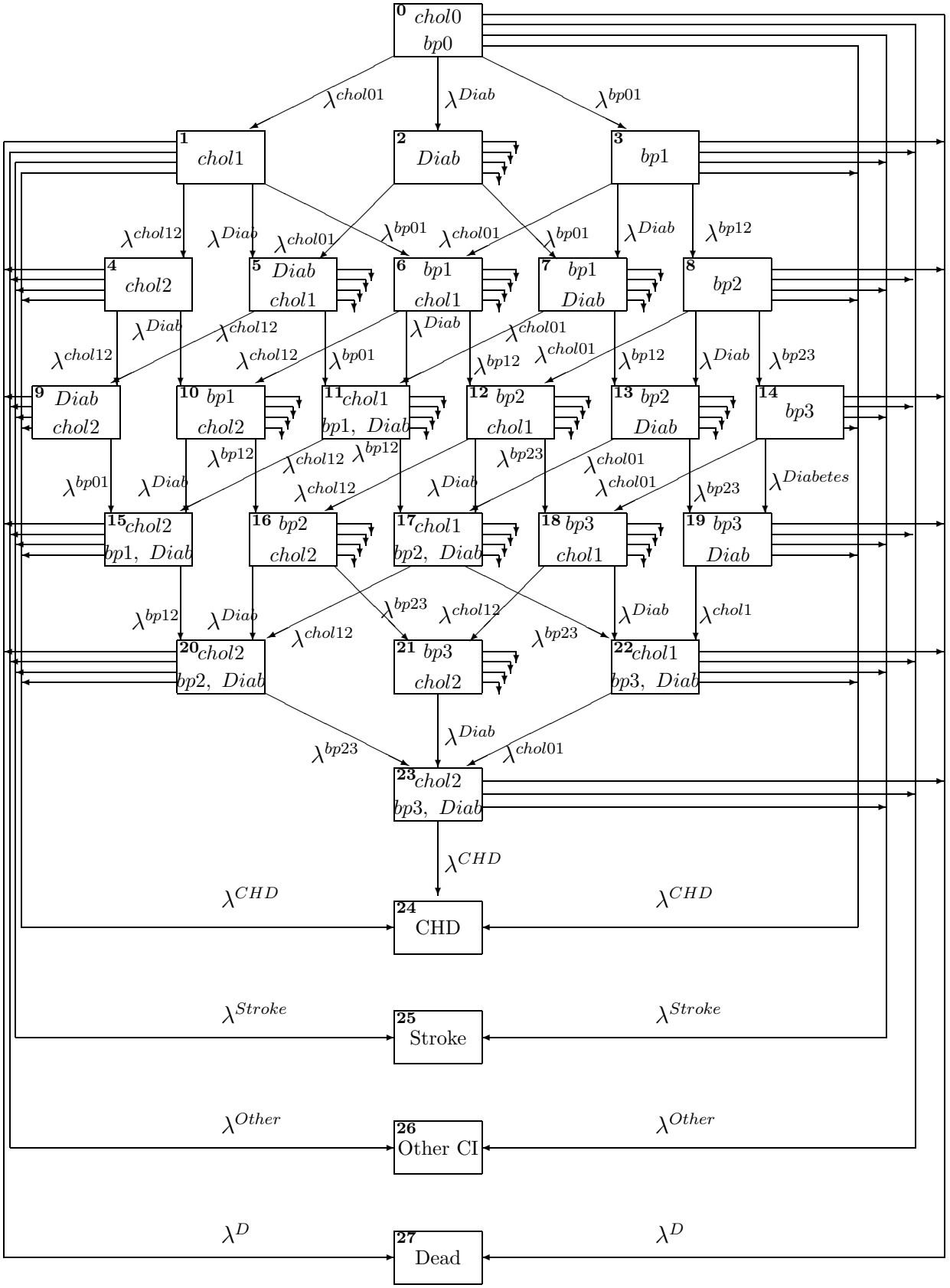


Figure 8: A CHD/Stroke model for Critical Illness Insurance.

risk factors that heighten the risk of MI are known, including age, sex, obesity (measured by body mass index (BMI)) blood pressure, levels of LDL cholesterol, diabetes, smoking, and social class. When we say these "heighten the risk of MI", we really mean that they stimulate the long process that leads to MI, the endpoint.

In Section 5 we saw that an actuarial model of the genetics of breast cancer had to include ovarian cancer as well, because the genes involved conferred extremely high risk of either cancer. Similarly, many of the risk factors for MI are also risk factors for stroke, so it is necessary to include both in a model. Macdonald, Waters & Wekwete (2003c, 2003d) proposed such a model, illustrated in part in Figure 8. The parameterisation of this model is described in the two papers above and here we will only mention that it was based on the Framingham heart study, a cohort study begun in 1949 and the source of much of the literature on CHD risk. Some features of the model are as follows:

(a) Three risk factors were chosen as categories that define twelve sub-populations; these were sex, smoking habit and BMI (three levels). Each sub-population is represented by the model in Figure 8, with appropriate intensities.

(b) Three risk factors were modelled as transitions between states in the model; these were diabetes, hypercholesterolaemia and high blood pressure. These had one, two and three levels respectively, so for example State 22 in Figure 8 is labelled "chol1, bp3, diab" and represents someone who has moderately high cholesterol levels, very high blood pressure and diabetes. A great deal of discussion with underwriters and clinicians led to the categories that were ultimately chosen.

(c) The model has four endpoints, one being a CI claim from a cause other than MI or stroke. This needed different parameters in different sub-populations where any of the 'other causes' shared risk factors with MI and stroke, for example smoking is a risk factor for lung cancer.

Genetic epidemiology is not yet at the stage where we can parameterise the intensities by genotype, but the model allows us to address a key question. When a 'gene for CHD' is discovered, it will not really mean that; it will mean a 'gene for a risk factor of CHD'. In other words, genotype will most likely be a risk factor for CHD (or stroke) because it

alters the level of one of the risk factors for the process leading up to CHD (or stroke). In the model, this means that genotype may alter the intensities in the upper part of Figure 8, but it is inlikely to alter directly λ^{CHD} or λ^{Stroke}.

Tables 11 and 12 show the essential difference between these two pathways for genetic risk to influence CHD and stroke. Table 11 shows the percentage extra CI premiums, for a male non-smoker, with normal BMI, aged 35 at entry with policy term 10 years, assuming that some gene is discovered that increases the risk of any of the three risk factors by a factor of 5. The left-hand column indicates risk factors that are already present at underwriting, which is the reason for the gaps in the right-hand part of the table. Relative risk of 5× is quite high for a multifactorial disorder (the U.K. Biobank project hopes to detect relative risks as low as 1.5× to 2×, depending on the exposures). Table 12 shows the effect of a gene that confers the same relative risk, but acts directly on the endpoints. The premiums are dramatically higher. Even if the relative risk is increased to exceedingly high levels, such as 50×, this conclusion holds: genes that act through risk factors do not result in exceptionally high extra premiums, genes that act on endpoints do. More examples can be found in Wekwete (2003). Note that these extra premiums take no account of any improvement in health that may be brought about by greater genetic knowledge of CHD and stroke, which is the whole point of the exercise from the medical community's point of view.

The usefulness of this model, and its conclusions, lies partly in the future, when information exists to allow its intensities to be parameterised with some degree of reliability. That may be some time away. However, it is entirely predictable that as genes are discovered that influence common disorders, we will see again and again a public outcry against insurance and the creation of a 'genetic underclass'. This work provides a good deal of evidence that this is not likely to happen.

7. A program of future research

Looking back to 1995 or so, when actuarial involvement in genetics and insurance began, we can see that that research has followed quite closely the emerging epidemiology (as it must) and that that has mostly been about the rare, late-onset single-gene

Table 11: Premium ratings for males, non-smokers, normal BMI aged 35 at entry with policy term 10 years, under hypothetical assumptions of genetic influence increasing the incidence of risk factors 5 times. Source: Macdonald, Waters & Wekwete (2003d).

	Premium rating factors with 5× the incidence rate of				
Risk factors	None	H'chol	H'tension	Type 1 Diabetes	Type 2 Diabetes
	%	%	%	%	%
No risk factors	0	+6	+14	+2	+5
Hypercholesterolaemia Cat 1	+4	+14	+19	+5	+9
Type 1 Diabetes	+247	+255	+265		
Type 2 Diabetes	+60	+68	+79		
Hypertension Cat 1	+6	+14	+33	+8	+12
Hypercholesterolaemia Cat 2	+33		+56	+35	+39
Hypercholesterolaemia Cat 1, Type 1 Diabetes	+251	+265	+271		
Hypercholesterolaemia Cat 1, Type 2 Diabetes	+65	+78	+85		
Hypercholesterolaemia Cat 1, Hypertension Cat 1	+11	+23	+39	+13	+16
Hypertension Cat 1, Type 1 Diabetes	+255	+264	+290		
Hypertension Cat 1, Type 2 Diabetes	+68	+78	+104		
Hypertension Cat 2	+43	+54	+64	+45	+49
Hypercholesterolaemia Cat 2, Type 1 Diabetes	+290		+319		
Hypercholesterolaemia Cat 2, Type 2 Diabetes	+103		+133		
Hypercholesterolaemia Cat 2, Hypertension Cat 1	+44		+86	+46	+50
Hypertension Cat 1, Hypercholesterolaemia Cat 1, Type 1 Diabetes	+260	+276	+298		
Hypertension Cat 1, Hypercholesterolaemia Cat 1, Type 2 Diabetes	+74	+90	+112		
Hypertension Cat 2, Hypercholesterolaemia Cat 1	+49	+69	+71	+51	+56
Hypertension Cat 2, Type 1 Diabetes	+303	+317	+330		
Hypertension Cat 2, Type 2 Diabetes	+116	+130	+144		
Hypertension Cat 3	+97	+113		+99	+105
Hypercholesterolaemia Cat 2, Hypertension Cat 1, Type 1 Diabetes	+304		+358		
Hypercholesterolaemia Cat 2, Hypertension Cat 1, Type 2 Diabetes	+117		+172		
Hypercholesterolaemia Cat 2, Hypertension Cat 2	+103		+134	+105	+111
Hypertension Cat 2, Hypercholesterolaemia Cat 1, Type 1 Diabetes	+311	+336	+340		
Hypertension Cat 2, Hypercholesterolaemia Cat 1, Type 2 Diabetes	+124	+150	+154		
Hypertension Cat 3, Hypercholesterolaemia Cat 1	+106	+134		+108	+114
Hypertension Cat 3, Type 1 Diabetes	+376	+396			
Hypertension Cat 3, Type 2 Diabetes	+189	+210			
Hypertension Cat 2, Hypercholesterolaemia Cat 2, Type 1 Diabetes	+381		+421		
Hypertension Cat 2, Hypercholesterolaemia Cat 2, Type 2 Diabetes	+195		+235		
Hypertension Cat 3, Hypercholesterolaemia Cat 2	+183			+186	+193
Hypertension Cat 3, Hypercholesterolaemia Cat 1, Type 1 Diabetes	+388	+423			
Hypertension Cat 3, Hypercholesterolaemia Cat 1, Type 2 Diabetes	+201	+237			

Table 12: Premium ratings for males, non-smokers, normal BMI aged 35 at entry with policy term 10 years, under hypothetical assumptions of genetic influence increasing the incidence of CHD and stroke 5 times. Source: Macdonald, Waters & Wekwete (2003d).

	Premium rating factors with 5× the incidence rate of						
Risk factors	None	CHD	Stroke	CHD modified by the presence of			
				H'chol	H'tension	Type 1 Diab	Type 2 Diab
	%	%	%	%	%	%	%
No risk factors	0	+192	+30	+32	+35	+1	+6
Hypercholesterolaemia Cat 1	+4	+210	+33	+210	+42	+4	+10
Type 1 Diabetes	+247	+497	+290	+287	+292	+497	
Type 2 Diabetes	+60	+311	+103	+101	+106		+311
Hypertension Cat 1	+6	+223	+37	+43	+223	+7	+13
Hypercholesterolaemia Cat 2	+33	+357	+63	+357	+92	+34	+43
Hypercholesterolaemia Cat 1, Type 1 Diabetes	+251	+519	+295	+519	+301	+519	
Hypercholesterolaemia Cat 1, Type 2 Diabetes	+65	+333	+108	+333	+114		+333
Hypercholesterolaemia Cat 1, Hypertension Cat 1	+11	+243	+41	+243	+243	+11	+18
Hypertension Cat 1, Type 1 Diabetes	+255	+536	+300	+302	+536	+536	
Hypertension Cat 1, Type 2 Diabetes	+68	+350	+114	+116	+350		+350
Hypertension Cat 2	+43	+393	+82	+100	+393	+44	+53
Hypercholesterolaemia Cat 2, Type 1 Diabetes	+290	+711	+333	+711	+364	+711	
Hypercholesterolaemia Cat 2, Type 2 Diabetes	+103	+525	+147	+525	+178		+525
Hypercholesterolaemia Cat 2, Hypertension Cat 1	+44	+407	+75	+407	+407	+45	+55
Hypertension Cat 1, Hypercholesterolaemia Cat 1, Type 1 Diab.	+260	+561	+305	+561	+561	+561	
Hypertension Cat 1, Hypercholesterolaemia Cat 1, Type 2 Diab.	+74	+375	+119	+375	+375		+375
Hypertension Cat 2, Hypercholesterolaemia Cat 1	+49	+424	+89	+424	+424	+50	+60
Hypertension Cat 2, Type 1 Diabetes	+303	+758	+360	+376	+758	+758	
Hypertension Cat 2, Type 2 Diabetes	+116	+572	+174	+190	+572		+572
Hypertension Cat 3	+97	+597	+203	+176	+597	+98	+111
Hypercholesterolaemia Cat 2, Hypertension Cat 1, Type 1 Diab.	+304	+774	+349	+774	+774	+774	
Hypercholesterolaemia Cat 2, Hypertension Cat 1, Type 2 Diab.	+117	+588	+162	+588	+588		+588
Hypercholesterolaemia Cat 2, Hypertension Cat 2	+103	+691	+143	+691	+691	+105	+120
Hypertension Cat 2, Hypercholesterolaemia Cat 1, Type 1 Diab.	+311	+797	+369	+797	+797	+797	
Hypertension Cat 2, Hypercholesterolaemia Cat 1, Type 2 Diab.	+124	+611	+182	+611	+611		+611
Hypertension Cat 3 Hypercholesterolaemia Cat 1	+106	+640	+212	+640	+640	+107	+121
Hypertension Cat 3, Type 1 Diabetes	+376	+1027	+531	+477	+1027	+1027	
Hypertension Cat 3, Type 2 Diabetes	+189	+840	+344	+291	+840		+840
Hypertension Cat 2, Hypercholesterolaemia Cat 2, Type 1 Diab.	+381	+1144	+439	+1144	+1144	+1144	
Hypertension Cat 2, Hypercholesterolaemia Cat 2, Type 2 Diab.	+195	+958	+252	+958	+958		+958
Hypertension Cat 3, Hypercholesterolaemia Cat 2	+183	+1024	+289	+1024	+1024	+185	+206
Hypertension Cat 3, Hypercholesterolaemia Cat 1, Type 1 Diab.	+388	+1080	+542	+1080	+1080	+1080	
Hypertension Cat 3, Hypercholesterolaemia Cat 1, Type 2 Diab.	+201	+894	+356	+894	+894		+894
Hypertension Cat 3, Hypercholesterolaemia Cat 2, Type 1 Diab.	+489	+1580	+643	+1580	+1580	+1580	
Hypertension Cat 3, Hypercholesterolaemia Cat 2, Type 2 Diab.	+302	+1394	+456	+1394	+1394		+1394

disorders. Scientifically, this is only a modest revolution, since family medical history has characterised these disorders for a long time. Politically, it has been more significant, because public concerns about genetics in general have pushed genetic information to the front of the long-running debate about fair and unfair discrimination. By accident, it is becoming a test case for the development of 'evidence-based underwriting'.

Medical research into multifactorial disorders is not very advanced by comparison, but that is where advances may be expected in future. Doubtless, each headline discovery will be followed by fresh allegations of a 'new genetic underclass', and maybe will fall within the remit of GAIC and the HGCA, so the need for actuarial research is unlikely to diminish. The following list suggests some questions, that may or may not turn out to be the key to the next ten years of actuarial effort in this area, and to the ability of insurance industries to negotiate viable underwriting practices with their respective governments.

- **Question 1.** Heterogeneity: what do we do regard as the 'same' genetic disorder? Do mutations in APKD1 and APKD2 cause the same disease or different diseases? Is every number of CAG repeats in HD a separate disease?

- **Question 2.** Data limitations: how should underwriting be allowed to proceed when there is evidence of significant risk but inadequate data to provide quantitative risk estimates? Or, where genetic epidemiologists simply have not yet got round to providing them?

- **Question 3.** Multifactorial disorders: what will the molecular genetics actually tell us about risks during a human lifetime? In this paper we have suggested a multiple-state model for CHD and stroke, which extends work on single-gene disorders, but other models may also be appropriate for risk factors that are measured on a continuous scale. What will be the actuarial models that are best for handling multifactorial disorders?

- **Question 4.** Family history: can we disentangle genetics and the environment? To the extent that family history is genetic information, it may fall under the terms of any moratorium (as already happens for Mendelian disorders in some countries).

What are the arguments for distinguishing between different categories of family history for underwriting use?

- **Question 5.** CI insurance: is it viable at all? How can any form of CI insurance handle the kind of early diagnosis and better treatments that genetics might eventually bring?

- **Question 6.** Adverse selection: to date, all useful models of adverse selection rely on hypotheses about how people's choices to buy insurance or to have a genetic test influence each other. Economics models based on equilibrium arguments do exist but are even more abstract than actuarial models. However, the threat of adverse selection is still the major reason why policy-makers might hesitate to legislate. We need much better empirical studies on adverse selection and people's choices.

- **Question 7.** Medical expenses and pharmacogenetics: insurance is not the only financial issue arising from genetics. How will care costs, from acute to long-term, be affected by genetics? Models developed to address insurance questions might be well able to address these questions too.

- **Question 8.** Longevity: so far most research has concentrated on single-gene disorders, where longevity is not an issue. But there is mounting evidence of a genetic contribution to longevity; what will be the impact on pensions and long-term care?

- **Question 9.** Product design: can we design insurance products that hedge or offset genetic risks? An obvious example is a combination of a pension and a long-term care insurance policy. What else is possible?

We end with the strong suggestion that the 2004 ASTIN Colloquium in Bergen, Norway, would be the ideal time and place to present work on these and related issues.

References

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

American Academy of Actuaries (1998). *Genetic information and voluntary life insurance.* Issue Brief.

American Academy of Actuaries (2002). *The ine of genetic information in disability income and long-term care insurance.* Issue Brief.

ALRC (2001). *Issues Paper No. 26. Protection of human genetic information.* Australian Law Reform Commission (www.alrc.gov.au).

ALRC (2002). *Discussion Paper No. 66. Protection of human genetic information.* Australian Law Reform Commission (www.alrc.gov.au).

ALRC (2003). *Report No. 96. Essentially yours: The protection of human genetic information in Australia.* Australian Law Reform Commission (www.alrc.gov.au).

Bartram, C., Beckmann, J., Breyer. F., Fey, G., Fontasch, C., Irrgang, B., Taupitz, J., Seel, K. & Thiele, F. (2000). *Humangenetische diagnostik, wissenschaftliche grundlagen und gesellschaftliche konsequenzen.* Springer-Verlag, Berlin.

Breyer, F. (2001). *Optionen für dir regulierung von gentests in versicherungswesen — ökonomische bewertung und ausländische erfahrungen.* Cologne Re Seminar über der lebens- und krankenversicherung.

Brinkman, R., Mezei, M., Theilmann, J., Almqvist, E. & Hayden, M. (1997). The likelihood of being sffected with Huntington disease by a particular age, for a specific CAG size. *American Journal of Human Genetics*, **60**, 1202–1210.

Daykin. C.D., Akers, D.A., Macdonald, A.S., McGleenan, T., Paul, D. & Turvey, P.J. (2003). Genetics and insurance — some social policy issues. *To appear in British Actuarial Journal.*

Dicke, A. (2002). Perception vs. reality. Life insurance and genetic testing. *Contingencies*, November/December 2002, 32–36.

Doble, A. (2001). *Genetics in society.* Institute of Actuaries in Australia, Sydney.

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., Gaytther, 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. (2003). *Modelling the impact of genetic testing on insurance — early-onset Alzheimer's disease and other single-gene disorders.* Ph.D. thesis, Heriot-Watt University, Edinburgh.

Gui, E.H. & Macdonald, A.S. (2002a). A Nelson-Aalen estimate of the incidence rates of early-onset Alzheimer's disease associated with the Presenilin-1 gene. *ASTIN Bulletin*, **32**, 1–42.

Gui, E.H. & Macdonald, A.S. (2002b). Early-onset Alzheimer's disease, critical illness insurance and life insurance. *Research Report No. 02/2, Genetics and Insurance Research Centre, Heriot-Watt University, Edinburgh.*

Gui, E.H. & Macdonald, A.S. (2003a). Should multiple endocrine neoplasia be an insurance problem or not?. *Research Report No. 03/2, Genetics and Insurance Research Centre, Heriot-Watt University.*

Gui, E.H. & Macdonald, A.S. (2003b). The incidence rates of early-onset Alzheimer's disease associated with the Presenilin-2 and Amyloid Precursor Protein genes. *Research Report No. 03/3, Genetics and Insurance Research Centre, Heriot-Watt University.*

Gutiérrez, M.C. & Macdonald, A.S. (2002a). Huntington's disease and insurance I: A model of Huntington's disease. *Research Report No. 02/3, Department of Actuarial Mathematics and Statistics, Heriot-Watt University.*

Gutiérrez, M.C. & Macdonald, A.S. (2002b). Huntington's disease and insurance II: Critical illness insurance and life insurance. *Research Report No. 02/4, Department of Actuarial Mathematics and Statistics, Heriot-Watt University.*

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

Harper, P.S. (1996). *Huntington's disease.* W.B. Saunders.

Hodge, S.E. (2002), Ascertainment, in *Biostatistical genetics and genetic epidemioogy*, ed. Elston, R., Olson, J. & Palmer, L., John Wiley

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.

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.

HGC (2002). *Inside information: Balancing interests in the use of personal genetic data.* The Human Genetics Commission, London.

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. (2003a). Moratoria on the use of genetic tests and family history for mortgage-related life insurance. *To appear in British Actuarial Journal.*

Macdonald, A.S. (2003b). Genetics and insurance: What have we learned so far?. *To appear in Scandinavian Actuarial Journal.*

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. (2003a). The genetics of breast and ovarian cancer I: A model of family history. *Scandinavian Actuarial Journal*, **2003**, 1–27.

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

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2003c). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting I: The model. *Swiss Re/Heriot-Watt Research Report No. 03/1, Heriot-Watt University, Edinburgh.*

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2003d). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting II: Applications. *Swiss Re/Heriot-Watt Research Report No. 03/2, Heriot-Watt University, Edinburgh.*

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.

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

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

Pokorski, R.J. & Ohlmer, U. (2000). Use of a Markov model to estimate long-term insured lives' mortality risk associated with BRCA1 and BRCA2 mutations. *North American Actuarial Journal*, **4**, 130–148.

Pritchard, D.J. (2002). *The genetics of Alzheimer's disease, modelling disability and adverse selection in the long-term care insurance market.* Ph.D. thesis, Heriot-Watt University, Edinburgh.

Regenauer, A. (2000). *Genetic testing and insurance — a global view.* Munich Reinsurance Company, Munich.

Smith, C. (1998). *Huntington's chorea: A mathematical model for life insurance.* 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.

Tan, K.W. (1997). *The financial impact of genetic testing on annuities.* M.Sc. dissertation, Heriot-Watt University, Edinburgh.

Thompson, E. (1993). Sampling and ascertainment in genetic epidemiology: A tutorial review. *Technical Report 243, Department of Statistics, University of Washington.*

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.

Warren, V., Brett, P., Macdonald, A.S., Plumb, R.H. & Read, A.P. (1999). *Genetic tests and future need for long-term care in the UK: report of a Work Group of the Continuing Care Conference Genetic Tests and Long-term Care Study Group.* Continuing Care Conference, London.

Wekwete, C.T. (2003). *Genetics and critical illness insurance underwriting: Models for breast cancer and ovarian cancer and for coronary heart disease and stroke.* Ph.D. thesis, Heriot-Watt University, Edinburgh.

Wesley, D. (2002). Primary cutaneous melanoma — SEER mortality follow-up and Kaplan-Meier table analysis. *Risk Insights*, **6:1**, 18–21.