Sources of Adverse Selection in Insurance Markets with Genetic Information

Craig J. Adams

SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

HERIOT-WATT UNIVERSITY

DEPARTMENT OF ACTUARIAL MATHEMATICS AND STATISTICS,
SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES.

December 16, 2013

Abstract

In this thesis we quantify costs of adverse selection in insurance markets where there are multiple sources of adverse selection. We aim to find the relative impact of genetic information as one of these sources.

Using new data on the effects of components of a polygenic model of breast cancer, we model adverse selection in a critical illness insurance market. We confirm the results of a previous study, which used a simpler polygene model without details of particular genes, that polygenes pose a greater source of adverse selection risk than the major genes (BRCA1 and BRCA2).

In a start-up market for long-term insurance, we model the progression of adverse selection costs over time, where premiums are repriced to adapt to the information the insurer gains about its business mix from its claims experience. In a U.K. setting we find the greatest costs of adverse selection come from a hypothetical intermediate stage of dementia progression which is not visible to an insurer, while testing of the APOE gene poses very little risk. We find the U.K. government's proposed cap on care liability has very little impact on adverse selection costs, as it benefits a very small proportion of people.

Acknowledgements

Contents

List of Figures

List of Tables

Introduction

In this thesis we are concerned with measuring costs of adverse selection in insurance markets where there are multiple sources of adverse selection. The aim of the thesis is to estimate the relative impact of genetic information as one of these sources.

We will use multiple state Markov models to represent the insurance markets of interest (critical illness insurance and long-term care insurance) with states indicating health status, whether genetic test has been taken, presence of a family history and whether the life is insured. These will be parameterised in part using transition intensities from relevant previous studies, as well as making use of available data from prospective cohort studies to fit transition intensities ourselves. Other transition intensities depend on a complex pattern of genetic inheritance, to estimate these we will simulate the future life histories of a large sample of lives. For the purpose of illustrating the relative impact of sources of adverse selection, we do not consider the compatibility of data to be a concern. The simulated samples might diverge slightly from the populations we attempt to model, but the transition intensities we will use provide us a baseline for modelling health. From this baseline, we can observe the order of magnitude of the adverse selection costs from each source and allow us to understand how they interact. As with any model, careful consideration should be given over the appropriateness before applying our models to any other purpose.

Chapter 1 gives a background in some of the key concepts which we will use throughout the work. Firstly, we describe terms relating to genetics which will be useful to greater understand the work of this thesis. We next describe how adverse selection arises and review all the available literature on the costs of adverse selection due to genetics and on the evidence for its occurrence in insurance markets (or lack thereof). A recent ruling by the European Court of Justice regarding sex discrimination in the insurance industry, has implications to our work. We close this chapter with a brief outline of this ruling.

In Chapter 2 we consider a polygenic model of breast cancer, using known gene data, and use this to outline a model of critical illness insurance in order to calculate a cost of adverse selection. The multiple sources of adverse selection in this market are both related to genetics: two major genes whose mutations are rare but confer a very high risk of developing breast or ovarian cancer; and a polygenic component,

where each of a large number of 'polygenes' increases or decreases risk of developing breast cancer a small amount but their variants are common among the population.

Common to all previous studies assessing the cost of adverse selection (including our work of Chapter 2), has been the assumption of an established market, *i.e.* the adverse selectors have been buying insurance at that rate for such a period that premiums have already absorbed it. Their analyses involve calculating the percentage difference between premiums in a market with adverse selection and one without adverse selection. They can shed no light on how the premiums would get to this stage over time and what losses might be incurred in the process.

In Chapter 3 we take the modelling further by outlining a multiple state Markov model for a start-up market of long-term care insurance. With this model, we explicitly show the progression of adverse selection costs using the development of information that an insurer would gain from analysing the claims history of its existing business, to reprice premiums for new business. In long-term care insurance, a major cause of claim is associated with Alzheimer's disease, which has a genetic component, leading to genetics as a potential source of adverse selection. We incorporate adverse selection from sources other than genetic information, by including states of health in which the probability of reaching a claim is high and we assume the insurer cannot identify lives in these states through underwriting. To overcome the complication of insurance benefit amounts which depend on the value of previous benefit payments, we develop a simulation approach of estimating the expected present values of insurance benefits and premium payments.

In Chapter 4 we apply this long-term care insurance model to a United Kingdom setting. We parameterise benefits based on the cost of care provision in the U.K. and include a government proposal to limit the individual's liability for their care costs. To assess their impact over time, we calculate adverse selection costs under various scenarios in which different sources of adverse selection are allowed to influence buying behaviour.

Finally, in Chapter 5 we give the conclusions this work leads to and suggest where further work may be useful in light of the direction of research in the genetics field.

Chapter 1

Background

In this chapter we give some background and review some of the available literature on various concepts that will be used further in this thesis.

1.1 Genetics

Since this thesis is concerned with the use of genetic information in insurance markets, we start by giving an overview of genetics.

To introduce genetics requires an introduction to what genes are made of: deoxyribonucleic acid or DNA. Most living organisms contain the molecule DNA (the exception being RNA viruses) and use it to code genetic information. The structure of DNA is of two strands of polymers in a double-helix, joined by hydrogen bonds between the pairing of bases: adenine with thymine and cytosine with guanine (Watson and Crick, 1953).

Humans have 3.9×10^9 base pairs (Pasternak, 1999) and in most DNA containing cells these are arranged into 23 pairs of chromosomes in the cell's nucleus. The sex cells, or gametes, contain only one of each chromosome allowing one chromosome of each pair to be received from the father and one from the mother. Mitochondrial DNA, a sequence of only 16,600 base pairs located in the mitochondria of a cell, is inherited exclusively from the mother.

A gene is a sequence of bases. These sequences are used to piece together proteins from their building blocks, amino acids — a set of three bases corresponds to a particular amino acid or to start/stop the sequence. A gene's position on a chromosome is referred to as its locus (plural loci). Only around 1.5% of human DNA codes protein synthesis (Sudbery, 2002).

When cells reproduce, DNA replicates in order to pass on a complete set of chromosomes to the child cell, identical to the chromosomes of the parent cell. At various points along each chromosome, DNA unwinds and the hydrogen bonds joining the individual strands, are broken. The double stranded complementary nature of DNA

bases allows each strand to act as a template for the other. An enzyme, known as DNA polymerase, reads the exposed bases and joins the corresponding base to create the new strand. Since this replication begins at various points, each segment needs to be joined to form a complete chromosome.

Meiosis, the process of producing gametes, involves genetic recombination whereby portions of the maternal chromosome and portions of the paternal chromosome are joined to produce a new chromosome with genes from each of their own parents — from the offspring's point of view, this means a chromosome from its father contains sequences from both paternal grandparents and similarly for the chromosome from its mother. The probability for recombination to occur at any given location is small, so genes which are close together are likely to stay together. Whereas, genes which are further apart or are on a different chromosome, can be seen as being inherited independently.

Different forms of a gene are known as alleles. For each gene, a person can therefore be either homozygous (both alleles are identical) or heterozygous (two different alleles). Alleles can be associated with particular traits in a dominant (apparent in heterozygotes) or recessive (apparent only in homozygotes) fashion. The cause of this is related to the level of gene product being produced. Someone who is heterozygous for a particular trait is able to synthesise some level of functional gene product from their normal allele, the other variant they carry may result in overproduction, underproduction or synthesis of a different product. In the case of recessive traits, the level of functional gene product produced by heterozygotes is adequate for the normal trait to be shown. However, for dominant traits, the level is inadequate by heterozygotes and a variant trait is shown although there may be a delay until later in life while the lack/excess of gene product builds up to a point where the trait is shown. Someone who is homozygous for a recessive trait will show it because they have no normal version of the gene to produce sufficient functional gene product. The traits which are shown are known as the phenotype, whereas the versions of the associated genes is the genotype.

Actuaries are usually concerned with diseases which act dominantly rather than recessive. This is because recessive gene disorders are usually present from birth or childhood *e.g.* cystic fibrosis, whereas dominant gene disorders might occur later in life. So an applicant for insurance may not yet be experiencing any visible symptoms to tell the insurer that they are (at least) at risk of getting the disease.

Gene variations which have a frequency in the population of more than 1% are referred to as polymorphisms. In particular, single-nucleotide polymorphisms (SNPs) are polymorphisms with a difference in a single base. Errors in the DNA replication process occur regularly and are usually repaired quickly (De Bont and van Larebeke, 2004). However, they can persist and exist as mutations. To be passed on to offspring,

a mutation must be present in the sex cells.

Mutations of genes coding for proteins do not necessarily cause a change in phenotype. The probability that a genetic disease develops in a carrier of the associated version of a gene is its penetrance. Where this probability is very high, the gene is said to be deterministic and it is fully penetrant. However, if it gives a higher risk of the disease but the symptoms might never show, the gene is a susceptibility gene and has incomplete penetrance.

Penetrance of a gene can be regulated by factors other than the gene itself. The epigenetic processes, DNA-methylation and histone modification, act to either 'switch on' or 'switch off' a gene without altering the genetic sequence. These processes can be influenced by environmental factors including, for instance, a high-fat diet or cigarette smoking (Aguilera et al., 2010).

Hence, a fully penetrant, dominant single-gene disorder will be displayed by the carrier, whereas a recessive version may be masked for generations by the dominant allele. This was first observed by Gregor Mendel, giving rise to the name Mendelian inheritance.

In a polygenic disorder, multiple genes interact and each contributes to a change in risk of disease. This creates a more complex pattern of inheritance which can result in clusters of disease in families, but is not Mendelian.

To uncover the genes associated with diseases, geneticists have historically used linkage analysis to narrow down the area in the genome to look at in more detail. This involves testing all chromosomes of family members for markers (making use of the tendency of genes which are close together to stay together described above) which are carried by sufferers but not among healthy family members. This technique led to the discovery of the role of gene locations with rare but high-risk disease associated alleles, *e.g.* BRCA1 and BRCA2, associated with breast cancer; CFTR, which causes cystic fibrosis; and the HD gene, which causes Huntington's disease (Bailey-Wilson and Wilson, 2011). More recently, as technology has advanced and costs have reduced, genome-wide association studies (GWAS) have been used to analyse all genes at high resolution. These analyses have aimed to find loci related to complex, non-Mendelian disorders, by the use of thousands of unrelated participants.

Once the gene responsible for a disease is located, epidemiologists study its development. However for a rare disorder, sampling at random from the population is unlikely to include sufficient carriers. Instead, it is common to study the families of those known to be affected. This will mean oversampling from a subgroup which is not necessarily representative — it excludes families with carriers who have not shown symptoms. This ascertainment bias can therefore result in overestimates of the parameters of interest. Burton et al. (2000) describes this in greater detail.

For further details of human genetics, good textbooks (and also the source of much

of this introduction) are Pasternak (1999) and Sudbery (2002).

In our modelling we will make great use of the term *relative risk*, particularly in the context of gene variants. This is a means of expressing the transition intensity of an event for one cohort, relative to some baseline cohort:

$$\mu_a = \mu_{baseline} \times RR_a, \tag{1.1}$$

where μ_a and $\mu_{baseline}$ are the transition intensities for cohort a and the baseline cohort respectively, and RR_a is the relative risk for cohort a, relative to the baseline cohort. Using this representation of genetic risk, relative risks will easily into our chosen modelling framework (multiple state Markov models), allowing us to estimate transition intensities for different groups.

1.2   Adverse Selection

When an insurance company sets its premium rates, it uses a set of assumptions about the future mortality and morbidity of the lives it expects to buy the product. The population is not homogeneous however, and this pricing basis will produce premiums that some will see as cheap because they consider themselves more likely to claim than the rest of the population; while others will consider them expensive if they do not think they will get much return from the contract. Underwriting serves to alter the basic premium and offer more appropriate premiums where necessary, *e.g.* lives who consume large amounts of alcohol might be charged a rating of say +50% on a critical illness product; enhanced annuities give a higher rate of annuity but these are only available to lives in ill-health who would otherwise not gain much from a standard annuity; and so on.

If underwriters are limited in their ability to identify the heterogeneity of a population, lives who think they are likely to claim will see greater value from the contract and may be more likely to buy insurance. Conversely, lives who think they are unlikely to claim may be less likely to buy the product. This we refer to as adverse selection but we recognise that it is simply rational decision making on the part of the proposer — unlike the similar concept of non-disclosure, whereby the proposer fraudulently withholds information which the insurer is entitled to use. If adverse selection occurs in the market then the average risk of the pool will be greater than that assumed, the premiums will be insufficient to cover benefits and a loss will occur on the business.

More formally, we define *adverse selection* as the decision to buy insurance based on an assessment of one's own probability to claim to be higher than the market

average, implied by the premium rates on offer when the insurer is unable to make a full assessment.

This scenario relies on *asymmetric information* (or *information asymmetry*) which we define as the case where two parties hold differing levels of information which allows one party a strategic advantage.

A similar, but somewhat different, concept to adverse selection is that of *moral hazard*. Moral hazard occurs after the insurance contract has commenced and is defined as being where the insured is not disincentivised from taking risks because their insurance covers the potential resulting loss.

Adverse selection relies on information asymmetry, however economists largely ignored the impact of asymmetric information and its influence on decisions, assuming perfect information by all parties, until the seminal work of Akerlof (1970) and Rothschild and Stiglitz (1976). They showed this asymmetry leads to inefficiencies in the insurance markets.

Akerlof (1970) discusses the impact of low quality goods (or 'lemons' as he refers to them from his analogy to used car sales) driving out the high quality goods to such a point where no market for the latter exists at all. In an insurance market, when only the lemons (the lives with a high probability of claim) remain, the high premiums required could make the market unviable.

Rothschild and Stiglitz (1976) explored the nature of information asymmetry further in an insurance market with competition and illustrated using indifference curves, that pooling equilibria (where all are offered the same contract) cannot exist in their model. Additionally, constructing contracts which appeal differently to the different risk cohorts (by varying coverage size and price) might also not produce an equilibrium. They argue that in withholding the information, high-risk lives are no better off but low-risk lives are worse off and suggest it is best for the system if high-risk lives reveal what they know. To understand how pooling equilibria do exist in practice, Allard et al. (1997) added distribution costs to the model and found that economies of scale cause pooling equilibria to always exist when costs are large enough.

1.2.1 Genetic Adverse Selection

Underwriting where there is a sound statistical basis to suggest a life is more or less likely to claim is often acceptable. Indeed there are clauses written into anti-discrimination laws in the U.K. that permit difference in terms if it "is done by reference to information that is both relevant to the assessment of the risk to be insured and from a source on which it is reasonable to rely"[1]. However, there are

---

[1]Equality Act 2010, sch 3 para 21(1)(b)

certain pieces of information about an individual that are contentious.

In the United Kingdom there is a moratorium on the use of genetic test results in insurance underwriting, while some other jurisdictions, for instance the United States of America have the prohibition enshrined in law and Sweden extends this to the medical history of family members; whereas, in Australia there is a requirement to divulge test results. Under the terms of the U.K.'s moratorium, an insurer may use the results of a predictive genetic test only if it has been approved by the government and the sum assured is greater than a product specific limit:

- Life insurance — £500,000;

- Critical illness insurance — £300,000;

- Income protection — £30,000 pa.

At the time of writing, the only approved test is for Huntington's disease and only for life insurance.

Since this moratorium permits individuals to withhold genetic test results, there is an asymmetry in the information known to the two parties, and the individual is better able to assess their likelihood of claiming. Genetic testing is not currently a common occurrence, some tests are available through the National Health Service, *e.g.* for breast cancer genes, BRCA1 and BRCA2, but these are offered if there is a high probability of a mutation being present. Hence, the volume of asymmetric information will be relatively low. However, it is plausible that testing will become a more frequent as costs decrease and preventative medical treatments are developed.

Much debate over the ethics and social outcomes of underwriters having access to test results has ensued over the past two decades. On one side, the insurance industry argued that it was fair to the other policyholders that those who bring extra risk to the pool pay for this increased cost and that "genetic information, where it is actuarially relevant, is little different from other forms of predictive healthcare information" (Daykin et al., 2003). However opponents to its use see their genes as the most personal of information and want their privacy protected. They fear the creation of a genetic underclass who are uninsurable because of what they view as discrimination.

Before much research had been conducted, the press had hyperbolic reactionary quotes such as, "Non-disclosure of genetic-test results could spell the end of the life-insurance market,"[2] and headlines including, "Fears raised over genetic tests"[3]. In contrast to the extreme statements, the U.K. government has not legislated on the

---

[2]Attributed to Achim Wambach, University of Munich, by The Economist magazine on the 19th of October, 2000 — `http://www.economist.com/node/398173`

[3]The Observer on the 22nd of October, 2000 — `http://www.guardian.co.uk/money/2000/oct/22/personalfinancenews.observercashsection2`

subject, but instead set up advisory bodies who would issue advice based on available evidence: the Human Genetics Advisory Commission in 1996 which became the Human Genetics Commission (HGC) in 2000. The result is the agreement with the Association of British Insurers (ABI) to create and implement the moratorium on the use of genetic testing set out at the start of this section which came into existence in 2001 for 5 years (and renewed three times since, currently until 2017) until further research could be performed.

The genetics and insurance committee (GAIC) was given the task of approving genetic tests under the moratorium until it was disbanded in 2009. Although only one test went through the full approval process (others were submitted but GAIC was suspended and re-established before completion), under advisement from geneticist, Professor Sandy Raeburn, the ABI regarded eight genetic disorders being significant to insurance business:

- Breast and ovarian cancers (of the hereditary form);

- Early onset Alzheimer's disease;

- Familial adenomatous polyposis;

- Hereditary motor and sensory neuropathy;

- Huntington's disease;

- Multiple endocrine neoplasia;

- Myotonic dystrophy;

- Adult polycystic kidney disease, although this was later dropped from the list due to its testing method (discussed below).

To feed into the debate, actuarial research has aimed to quantify the impact on insurance business of not being able to underwrite when genetic information (with the possible inclusion of family histories) is withheld from the insurer.

Costs of Genetic Adverse Selection

The history of quantitative research in genetics and insurance begins with Macdonald (1997). This study used a 'top-down' approach to model all genetic disorders in a life insurance market: extreme assumptions are made regarding the frequency of genetic disorders, their impact on mortality and the rate of testing. The moments of the financial impact of not being able to use the results of genetic tests are calculated — in this case by solving Norberg's equations (Norberg, 1995) — and where these are negligible, it can be concluded that adverse selection from genetic testing is not a

concern. He found that adverse selection in the form of high risk lives buying larger sums assured posed a greater threat than those lives merely increasing their rate of purchase. Indeed, there was an approximately linear relation between the size of benefit purchased by adverse selectors and the size of loss, as a percentage of baseline costs.

Since the epidemiological consequences often take some time to emerge after the discovery of the role of genes in particular diseases, the simple assumptions of this method are an advantage. However, its ability to draw conclusions when the adverse selection costs are non-negligible, is limited. In this case, a 'bottom-up' approach — modelling particular genetic disorders with realistic assumptions — is more appropriate.

The first such bottom-up modelling was that of Lemaire et al. (2000). They calculated the relative prices of term assurance for women with a family history of BC or OC (which itself can be argued to be genetic information) and for women with mutations in either the BRCA1 or BRCA2 genes which are known to cause BC and OC. The epidemiology of BRCA1 and BRCA2 mutations available to them at this time was limited to estimates of the penetrance and a normal distribution for age of onset for a non-specific BRCA mutation.

In the follow-up to this work, Subramanian et al. (1999) considered the insurance purchase behaviour for lives before and after a test for BRCA1 or BRCA2 mutations and used this to calculate costs of adverse selection — the percentage by which premiums calculated assuming no adverse selection should be increased in the presence of adverse buying behaviour. Set in a Markov environment, this model allowed the possibility of buying varying sums assured at an increased probability after a positive test; and a possibly lower sum assured at a decreased probability of purchase after a negative test. Additionally the probability of lapsing the policy was lower after a positive test. When pricing in knowledge of any family history, these adverse selection costs were minimal. However, when family history could no longer be used as a rating factor, adverse selection costs were substantial.

Possibly because BC is the most prevalent form of cancer in the world (Parkin et al., 2001), a large body of research on the disease exists. This has yielded a progression of epidemiology which has been followed by a series of actuarial studies involving BC since the first work by Lemaire et al. (2000). We list the key milestones of this research here:

- The epidemiology available to Macdonald et al. (2003b) allowed them to calculate stand-alone critical illness premiums specific to those with either a BRCA1 or BRCA2 mutation respectively. Subsequently they calculated the adverse selection cost from lives testing positive for either of these buying insurance at an increased rate.

- Gui et al. (2006) included the development of a family history of BC or OC among first degree relatives (mother/sisters) as a state in their models of life insurance and critical illness markets. This allowed them to consider testing to be done after a family history has emerged (which is more realistic) and to calculate adverse selection in situations where family history can and cannot be used as a rating factor.

- Antoniou et al. (2002) proposed a hypothetical polygenic model of BC risk without epidemiology of particular genes. Macdonald and McIvor (2006, 2009) used this to calculate family history ratings and adverse selection costs in a critical illness insurance market.

- Viswanathan et al. (2007) used an elastic demand model to create a dynamic response from lives, to the price being charged by the insurer (which itself is dynamic based on how it thinks the population would respond). The underlying model of BC/OC was that of Subramanian et al. (1999). Elasticity of demand was calculated based on survey responses. They found price elasticity of demand to have only a minimal impact on adverse selection costs.

- Lu et al. (2011b) applied the semi-Markov model of progression of BC and OC from Lu et al. (2011a) (which included the major genes, BRCA1 and BRCA2 only) to an income protection insurance market and answered questions similar to those answered by the preceding studies. This model incorporated the varying treatments given to women with and without a BRCA1 or 2 mutation and the recurrence of their cancer later in life. During the period of any treatment (or during some other sickness), she would be unable to work.

In each of these studies the results were similar: premiums chargeable to lives with a high risk genotype would be very high and possibly beyond insurability limits, however in a market the size of the U.K., the costs of adverse selection would be negligible or small enough to be managed by the insurer. Chapter 2 of this thesis develops the polygenic model of BC risk in the advent of the development of actual epidemiology for some of the polygenes hypothecated by Antoniou et al. (2002).

Actuarial research involving other disorders has also been conducted, a significant volume of which has been done in the Genetics and Insurance Research Centre (GIRC).

The first bottom-up modelling performed by GIRC was on Alzheimer's disease (AD), a form of dementia which primarily affects older people, above 60 years old. It is known that the APOE gene is associated with AD (Saunders et al., 1993) however this is not the only component of the genetic risk. Macdonald and Pritchard (2001) calculated the adverse selection costs from lives who know they have a high risk

genotype for AD in a long-term care market. While the premiums appropriate for lives with the highest risk variant of APOE may be up to 40% higher than average premium, adverse selection costs were generally not significant. The only case where single-premiums would have to rise significantly is when the market is small, APOEϵ4 carriers (the high risk variant) have risks of developing AD at least as high as observed and are highly likely to buy insurance and there is extensive testing of APOE.

AD can also affect people under 60 years old, in a form known as early-onset Alzheimer's disease. The genes involved in early-onset AD are known to include APP, PSEN-1 and PSEN-2. Gui and Macdonald (2002) modelled critical illness and life insurance markets in a Markov and semi-Markov framework respectively — due to the survival after diagnosis being duration dependent, the life insurance market is semi-Markov. They considered a method whereby they paid out the value of the reserve upon transition into the early-onset AD state or the sum assured on death without early-onset AD. This is the same method referred to as 'instantly reinsuring' in Gui et al. (2006). They found similar conclusions to those of Macdonald and Pritchard (2001), that adverse selection costs are negligible unless the market is small and buying behaviour of high risk lives is 'extreme'.

Adult polycystic kidney disease (APKD) is a single gene disorder caused by mutations in one of at least two genes: APKD1 and APKD2. It can lead to end stage renal disease (kidney failure), thereby triggering a critical illness claim. Before DNA-based testing was available, at risk individuals (those with a family history of APKD) could be tested with ultrasound for signs of cysts, long before symptoms had developed. Since this does not involve direct testing of DNA or chromosomes, it was outside the narrow definition of a genetic test in the moratorium. Gutiérrez and Macdonald (2003) modelled a critical illness insurance market where adverse selection comes after a positive test result which is hidden from the insurer. They also considered the cases where negative results were also hidden and where family history rating was not permitted. Again, while premiums for those carrying an APKD mutation were large, the costs of adverse selection were not very high — around 5% increase to premiums in a small market when the moratorium is extended to family history.

With the development of reliable DNA-based tests for APKD, Gutiérrez and Macdonald (2007) revisited their previous work to perform their analysis with the 3 genetic subgroups (no mutations, APKD1 mutation and APKD2 mutation) now possible, as opposed to their previous 2 genetic subgroups (no mutation, mutation of either APKD1 or APKD2). Additionally, they assumed dialysis is available to those with ESRD and the possibility of a kidney transplant (with varying scarcity) to model the development post-ESRD to extend the model to a semi-Markov model of a life insurance market. One of the more interesting points of this work was that individuals with a family history could be charged lower premiums if they tested positive

for APKD2. However, as this would be divulging an 'adverse' test result, this would not be permitted under the moratorium and consequently the individual would not be able to benefit from this extra knowledge.

Huntington's disease (HD) is a neurological disease which leads to loss of cognitive function and death and whose sole cause is from mutations in the HD gene. These mutations are of the form of repetitions of the trinucleotide, CAG, in a particular region of the gene. Penetrance depends on the number of CAG repeats with a lower average age of onset as number of repeats increases: < 36 repeats is normal and won't develop HD; 36–39 repeats, the individual could develop HD; ≥ 40 repeats will cause HD. Gutiérrez and Macdonald (2002) modelled the onset and development of HD and they applied it to critical illness insurance and life insurance markets in Gutiérrez and Macdonald (2004). Although the positive test result for the HD gene is the only one which is permitted under the moratorium, the adverse selection costs calculated by Gutiérrez and Macdonald (2004) were negligible for life insurance and very small for critical illness insurance (where the test results are not usable).

MacCalman (2009) performed further analysis on HD to address an issue of reliability of insurance estimates when epidemiological data could be affected by ascertainment bias. She fitted a Normal distribution for age-at-onset which did not include the allowance of number of CAG repeats that Gutiérrez and Macdonald (2004) included. She confirmed the high premiums that would be necessary when genotype of HD is known to be adverse.

Lu et al. (2007) modelled hereditary nonpolyposis colorectal cancer (HNPCC), a form of cancer which is caused by mutations in any of MLH1, MSH2, MSH6, PMS1 and PMS2 (the DNA mismatch repair genes). Colorectal cancer is one of the most common forms of cancer and, like many other cancers, it can be sporadic or hereditary. The majority of HNPCC cases are a result of mutations in MLH1 and MSH2 so these are what they considered in their modelling of a critical illness insurance market. Premium rates for people with a mutation exceeded the common limit of insurability (extra premiums of 250%), while those with a family history of the disease could be charged an extra premium of 30–374% depending on entry age and term. However, the cost of adverse selection on genetic tests or family history was very small.

Although each of these studies has found that in isolation the adverse selection costs are small, they noted a caveat that it could mount up when considering a bigger picture, with all genetic disorders contributing. Macdonald and Yu (2011) brings together all of the major gene disorders discussed above (the polygenic model of BC is excluded), with the addition of myotonic dystrophy (MD). By piecing together the models for each disease, they model them all in the context of both critical illness and life insurance markets. In each market, the impact of a moratorium on genetic testing was minimal and on use of family history small enough that they would not

have 'any measurable effect.'

Economic Outcomes

The impact of adverse selection has also been measured in terms of welfare and market efficiency. Before genetic testing was the issue, tests for human immunodeficiency virus (HIV) were a concern in a similar manner. Doherty and Thistle (1996) found that when information was symmetric between the insurer and individual, the value of the information was negative so insurance discourages them from being tested. This is due to a risk-averse individual preferring to avoid the lottery they would be faced with around the high and low risk policies for which they would be charged. Obviously, a public health goal would not be to deter testing so this is an undesired outcome. However, if test results cannot be observed, there is a positive private value in getting tested but at a trade-off in loss of market efficiency since premiums would be increased to cover adverse selection.

A similar study was performed by Hoy and Polborn (2000) in the case of a life insurance market — where the loss to be covered is not clearly definable like it is in other markets. They created scenarios in which a new test could lead to either Pareto improvements or Pareto worsenings. They suggest that when only few people are tested, it is likely that prices will not change much but those being tested will benefit.

Hoy and Witt (2007) applied their results to the genetic testing of BRCA1 and BRCA2. In a simulated market of 10-year term assurances offered to women aged 35 to 39, they showed that market efficiency could suffer when a large proportion of women made use of genetic tests.

Evidence of Genetic Adverse Selection

Justification for the existence of genetic adverse selection can be found in some of the available literature but most studies do not find evidence to reject a hypothesis of no adverse selection in the market. Understandably from the rarity of mutations, empirical evidence for adverse selection based on genetic testing is limited. Some studies have been done, which we review below, but we will see that their results are not necessarily reliable.

Macdonald and Tapadar (2010) aimed to find how high average insurance premiums would have to go before low-risk lives would stop buying, thereby creating adverse selection, using utility theory. They argued that adverse selection would occur if the expected utility from insurance purchase was lower than if insurance was not purchased. Under their model, purchase was binary: the life bought insurance if this improved their expected utility, or they did not buy it. They fitted 4 models of

utility, for a genetic disorder with environmental interactions, and found no evidence that adverse selection would cause a serious risk.

Macdonald and McIvor (2009) performed a similar analysis, using the same utility functions as Macdonald and Tapadar (2010), with their own BC polygene model. Their findings were that very low-risk polygenotypes would not buy insurance under 2 of the utility models.

Both of these studies assumed rational behaviour on the part of the customer, without any allowance for emotion. However, knowledge that one's risk profile implies a higher than average probability of disease can create anxiety (Aktan-Collan et al., 2001), while Loewenstein (2000) outlines how this type of visceral factor can have 'important, but often unappreciated consequences for behaviour'.

The first study to use empirical data to assess the existence of adverse selection was Zick et al. (2000). They were concerned with how testing for BRCA1 might influence purchase of life insurance. Their 'at-risk' group comprised female members of kindred K2082, a large extended family in Utah and Idaho which is known to be at risk of having a particular BRCA1 mutation (Goldgar et al., 1994; Botkin et al., 1996). They were given genetic counseling before any decision to take a test and at the receipt of test results. The control group comprised of women who had previously been interviewed for having a family history of BC or OC. The same selection criteria — no missing data, no life insurance paid by employer and no personal history of cancer — was applied to both groups. Their statistical analyses showed that neither positive, nor negative test results, were significant factors in buying insurance giving insufficient evidence to reject the hypothesis of no adverse selection. They suggest reasons why there was not stronger evidence of adverse selection. These include that these women and the insurers have knowledge of their family history so a test 'serves to confirm' what they already suspected — their behaviour might have been impacted upon first suspicion but at the same time, the insurers would put up their premiums. Additionally, they suggest that a year's follow-up period may have been too short as the women who tested positive would have more pressing concerns than insurance, *e.g.* how to reduce their risk of cancer. As the authors acknowledged, most of the women in the study identified themselves as 'active members of the Church of Jesus Christ of Latter Day Saints', a very patriarchal religion which supports traditional gender roles. Since life insurance tends to be used as a means of protecting income, a group where women are often housewives might be expected to have lower life insurance cover. Putting these two points together could make generalising the results invalid.

Aktan-Collan et al. (2002) surveyed participants of a genetic testing programme for HNPCC one year after their test. Prior to testing, all participants were considered to have 50% chance of having a mutation based on confirmations of mutations in family members in a previous study. In a questionnaire the participants were asked about life

and health insurance before and after the test. Although they found no significant increase in purchases after the test which would have illustrated adverse selection from the test results, a significantly higher proportion of mutation carriers had health insurance before even having the test. They suggest respondents not wishing to admit that they had used the test results as means of making their decision to buy could be a possibility for this anomaly.

Armstrong et al. (2003) performed a retrospective cohort study on women who had received genetic counseling and/or a genetic test for BRCA1 or BRCA2 mutations at their university clinic. Their counseling involved up to 3 stages: general information about BRCA1 and BRCA2 testing and breast cancer risk; an individualised assessment of their breast cancer risk before testing; women who agree to testing receive their results with 'further counseling' to explain its implications. For this study, the participants were mailed questionnaires to ask about life insurance and their experiences in purchasing it. The results showed an association with the predicted level of breast cancer risk and increases in life coverage. Additionally, carriers of a BRCA1/2 mutation were significantly more likely to increase the life cover.

Zick et al. (2005) used the results of a randomised controlled trial that aimed to evaluate genetic education and counseling for adult children of AD patients. Questions about insurance were only a part among a wider whole in the evaluation. The control group were told of their AD risk based on sex and family history, while the other subjects were tested for at least one APOEε4 allele, the high risk variant, and advised based on this additional information and given counseling. Participants were followed for a year after the initial interview. In a bivariate analysis, the authors found the percentage of APOEε4 positive subjects who had changed their long-term care insurance coverage to be significant, as was the percentage who were considering changing their level. Taylor et al. (2010)'s research using a later iteration of the trial which went further in genotyping, agrees with this finding. However, in a multivariate analysis by Zick et al. (2005), which controlled for confounding factors, this significance was reduced, such that the p-value was 5.11% for a positive test result being a factor in those who had changed. The sample size was small and the authors stressed the need for a larger and more diverse sample.

Oster et al. (2010) compared ownership levels of long-term care insurance of lives at risk of HD with the general population. Their study followed a group of lives with a family history of HD who were to be given a genetic test, asking them at various intervals about their level of insurance coverage. They compared this with details of the Health and Retirement Survey (HRS) of individuals over 50 years old, which also includes a small number of individuals under 50 years old, because it asks about long-term care insurance ownership. Since HD has usually manifested by the time people are old, they excluded people older than 65 from the HRS dataset in

an attempt to address the mismatch. In comparing ownership after controlling for various factors including age, they found that the individuals who they were following had a significantly higher ownership than the general population. Additionally, they found a significant difference between ownership among those with a positive test result and those with a negative test result, implying adverse selection based on test results. However their design has flaws which could invalidate their results. They acknowledge that there is an issue with the sample size of tested individuals and that there is selection bias. However, they do not consider the effect of repeatedly asking individuals whether they have insurance to cover the potential costs of their care. Since this is not a product that is commonly purchased, particularly at the ages in the dataset, this could have the potential to affect a significant proportion of purchases.

Evidence of Adverse Selection — Other than Genetics

We now turn to examples of adverse selection being found without specific cause for its occurrence. If genetics is considered as merely one of many ways that individuals can measure themselves as likely or unlikely to claim under a policy then there may be no reason to assume that their decision making process would differ and we review a selection of work on more general adverse selection in insurance markets.

Health insurance in the United States is often provided as an employment benefit so commonly people are automatically enrolled in a group policy. Brown (1992) measured adverse selection in terms of the underbuying of lives who see themselves as low risk. Since automatic enrollment reduces adverse selection in the group market (low-risk lives would be unlikely to opt out as it is paid for by the employer), he suggests that the individual market would have higher potential adverse selection. He uses data from a national survey which collected details of insurance levels and opinions of health from 2,745 randomly selected families (2,515 with a group policy and 225 in the individual market). From this he estimated a measure for the quantity of insurance for each family as a factor of the insured's risk characteristics and design of the policy. To find the discrepancy between predicted and actual insurance purchase, he fitted a linear model which controlled for various variables expected to influence purchase. A response of good health in the survey was a significant factor in this model, hence he concluded the presence of adverse selection. He notes however, it may be possible that low-risk lives are off their demand curve (*i.e.* in receipt of more insurance than they would otherwise purchase) in the group market, hence the 'control' does not necessarily follow what behaviour would truly be, distorting results.

Finkelstein and Poterba (2004) analysed a U.K. insurance company's book of compulsory and voluntary annuity business, sold over a 17-year period, for evidence of adverse selection. They focused on three features of annuities: the initial annual

payment; degree of back-loading (from some form of escalation); and size of payments to the estate after death. Since these features would appeal differently according to how a person views their mortality (assuming the insurer has not got the information to make the pricing indifferent to mortality *i.e. asymmetric information*), they analyse the mortality experience with a statistical model. Their results show significantly lower mortality among lives who bought policies with a high degree of back-loading in both markets; in the voluntary market there was significantly higher mortality among those who bought policies with payments to the estate, however the in the compulsory market this was only the case among lives who bought a 10 year guarantee period. They argue that this suggests some unobserved factors of mortality is correlated to the choice of annuity and is consistent with models of adverse selection. They highlight that adverse selection is but one possibility and suggest that preference, for instance to leave a bequest, could be the causative factor if they are correlated with mortality even if the individual is unaware.

Finkelstein and McGarry (2006) investigated the nature of private information in a long-term care insurance market. They analysed results of the HRS, which in addition to asking about about long-term care cover, also asks participants about nursing home use and the expectation of moving to a nursing home among other questions. Their findings included a positive correlation between private information about level of risk of going into a nursing home and insurance coverage (adverse selection). However, in looking at the actual experience, they found nursing home care use not to be positively correlated with insurance coverage. They suggest the discrepancy is partly due to pessimism — individuals who actively try to improve health, thereby avoiding/delaying nursing care, are likely to overestimate their risk.

This pessimism could be considered an example of the inability to assess one's own risk described by Siegelman (2003). He views methods used for testing adverse selection as flawed because they rely on an individual to assess their own health which is inherently difficult. Indeed, he suggests an insurance company who receives a medical report about the individual will no doubt have a better ability to assess the risk. While he concedes that adverse selection does exist, he considers the impact of adverse selection on the markets to be the more important aspect for researchers.

Courbage and Roudaut (2008) analysed the French market for long-term care insurance using the second wave of the Survey of Health, Ageing and Retirement in Europe (Börsch-Supan et al., 2005), a database of over 22,000 Europeans over 50 years old with details including various influences on health and insurance ownership. Their aim was to understand the variables influencing demand for long-term care insurance in the French market, one of the largest for this type of insurance in the world. They found that individuals with high alcohol consumption or body mass index have a higher probability of owning long-term care insurance. Additionally,

those who self-reported their health to be bad had a higher probability of buying the insurance.

1.3   Sex Discrimination

Across most types of insurance, experience generally differs by sex. In some types, the difference reverses as age progresses *e.g.* critical illness morbidity is lower for men at young ages but higher after around age 45, when males start to have a higher risk of stroke and heart attack (see Chapter 2).

In the U.K., sex was permitted to be used as a rating factor under the Equality Act (2010) and previously the Sex Discrimination Act (1975) in a similar wording to that for disability mentioned above. Similar practice existed in many countries across the European Union and the wider world, as limited by local laws. In 1985, the insurance industry's interpretation of the law was tested in the U.K. legal system in a case regarding a woman being charged more for her permanent health insurance[4] with the court finding in favour of the insurance company. Thus clear biological differences (and statistical correlations in the untested case of car insurance) prevailed in allowing sex as a pricing factor.

This legal exemption to discriminate was brought to an end in March 2011, when the European Court of Justice ruled in a case brought initially before the Belgian courts by Belgian consumer organisation Test-Achats. The court's finding was that pricing with sex as a rating factor is contrary to the principles of the European Union and would not be permitted from 22nd December, 2012 (European Court of Justice, 2011).

Sex is a genetic difference, determined by the sex chromosomes: XX for females and XY for males, but it is not unknown to the insurer. The Court's ruling restricts insurers only for using sex in their pricing (premiums must be equal for males and females). When setting reserves or performing valuations the insurer is not prevented from using the information of a policyholder's sex.

We assume that insurers' responses will be to charge premiums based on aggregate risk on an assumed business mix. There is the potential that this could increase adverse selection. However, while insurers are able to collect detail of sex from the application (or at least infer to some extent from the name) the response time to adjust prices should restrict the potential adverse selection costs. The ability to determine the sex of the customer should also ensure capital positions are not subject to uncertainty of future benefits/premium due to the mix of sex.

---

[4]Pinder v The Friends Provident Life Office (County Court, 15.8.85) EOR5D

Chapter 2

Polygenes and Major Genes for Breast Cancer in a Critical Illness Market

2.1 Introduction

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

Rare inherited mutations in either of two genes, BRCA1 and BRCA2, confer greatly increased risk of BC and of ovarian cancer (OC). These are known as 'single genes' or 'major genes' because mutations in either are sufficient to increase the risk, in this case by affecting production of a DNA-repairing protein produced in the BC pathway (Tutt and Ashworth, 2002). Mutations come in many different forms, and although some are benign, others cause a very high risk of BC or OC. Epidemiology rarely goes to the level of particular mutation, so they are commonly treated as being homogeneous and it is only the presence of a mutation that determines the risk. Moreover, occurrences of mutations are more common within subgroups of the population, *e.g.* the frequency of BRCA1 185delT and BRCA2 6174delT among the American Ashkenazi Jewish population is about 1% each (Sudbery, 2002), compared to about 0.2% each for any mutation in either BRCA1 or BRCA2 based on Antoniou et al. (2002)'s prevalence estimates.

The implications for life, critical illness and income protection insurance of these genes have been extensively studied (Lemaire et al., 2000; Subramanian et al., 1999; Macdonald et al., 2003a,b; Gui et al., 2006; Viswanathan et al., 2007; Lu et al.,

2011a; Lu et al., 2011b). These studies came to the same general conclusions: high premiums for lives known to have mutations, however the overall impact of adverse selection when underwriters cannot use test results is minimal.

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

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

Macdonald and McIvor (2006, 2009) incorporated this polygenic model, along with models of BRCA1 and BRCA2 risk, into a Markov model of stand-alone critical illness (CI) insurance. They estimated premium ratings conditional on either knowing an individual's genotype (assuming genetic testing for the major genes and the polygene) or a family history of BC and OC. They found that the larger proportion of premium increases was the result of the polygene. However, they suggested that the heavy tails in the distribution of relative risks caused by there being only three genetic loci in the polygene could have skewed the results.

Seven single nucleotide polymorphisms (SNPs) that affect the onset of BC have subsequently been described (Easton et al., 2007; Hunter et al., 2007; Stacey et al., 2007; Cox et al., 2007). They are all candidates for the actual genetic loci in the polygene hypothesised by Antoniou et al. (2002). However, since in total they contribute about 3.6% of the heritable risk of BC (Easton et al., 2007) and the two major genes another 25%, it is highly probable that the number of genetic loci involved will be well over 100. We describe these developments in Section 2.2.1, and extend the findings to a plausible polygenic model accounting for 75% of the total heritable risk. We note that even since commencing work on this project using these seven SNPs, another eighteen loci have been identified with similar distributions, validating the assumptions we make in Section 2.2.2. The identification of the genetic loci that contribute to the risk of BC is progressing quickly and replication of results is already being sought for a further sixty-nine loci (Ghoussaini et al., 2012).

In Section 2.3 we estimate the effect on stand-alone CI premiums of this polygenic model, given a family history of BC or OC, along the lines of Macdonald and McIvor (2006). We simulate the life histories of women within families. We find the distribution of genetic risk within families that develop a family history of BC or OC, and families that do not.

The European Court of Justice in March 2011 removed sex as a permissible insurance underwriting factor, from December 2012. This has particularly interesting implications for the pricing of BC risk, because men and women carry genotypes and develop family histories equally, but the risk of developing BC overwhelmingly affects women (male BC exists but is very rare). Thus men may disclose information affecting BC risk while being at virtually no risk of developing it themselves. We adapt our model to this new European context in Section 2.4. Additionally, in Section 2.5, we adapt the model to the terms of the moratorium in force in the United Kingdom which limits the use of genetic information in underwriting. This extends the insurance market model of Macdonald and McIvor (2009) to potential adverse selection in the light of recent epidemiology and in the new legal environment.

The work in this chapter has been published in the Scandinavian Actuarial Journal (Adams et al., 2013).

2.1.1 Review of Previous Actuarial Models

In Chapter 1 we reviewed the findings of previous models of adverse selection across all of the genetic studies. Here, we review the modelling methodology in the previous actuarial studies briefly mentioned above, pertaining to BC and OC.

Lemaire et al. (2000)'s life insurance model was of the form of a double decrement for death before BC or BC development with a single decrement for death from any cause after BC. Their approach used one-year probabilities of decrement and an assumption of mid-year payment of benefit to calculate single premiums by summing over possible outcomes. This was done for lives with varying family history details (one or two first-degree relatives; BC or OC; and age of onset of affected relative) using the probabilities derived by Claus et al. (1994) for BC and Hartge et al. (1994) for OC. For BRCA mutation, they performed their modelling by use of penetrance and age at onset distribution of BC from Claus et al. (1994) and an estimated adjustment to population probabilities for OC based on the broad range of estimates in the medical literature.

Subramanian et al. (1999) set out a Markov model for the term assurance market. They calculated ratios between the forces of mortality for women with family history or BRCA mutation compared to the baseline mortality based on the model of Lemaire et al. (2000). Their market model worked as follows:

- Women started untested, either insured or uninsured;

- Family history during modelling and was set according to the scenario;

- Testing for BRCA mutation was at a constant rate, with the probability of positive result determined by family history;

- Test results would influence a woman's decision to purchase life insurance, change sum assured or lapse her policy.

In common with subsequent models, they calculated average single premiums by solving Norberg (1995)'s equations and compared to their baseline to calculate costs of adverse selection.

Macdonald et al. (2003a) used a Markov model for BC/OC to estimate the conditional probabilities of BRCA1 or BRCA2 mutations (epidemiology had developed allowing them to consider these separately), given at least partial knowledge of family history. They used a 'brute force' method, summing over the possible family structures and the associated probabilities. Transition intensities for the development of BC and OC with a BRCA1 or BRCA2 mutation were fitted to Ford et al. (1998) and the distribution of the mutations came from Parmigiani et al. (1998). Macdonald et al. (2003b) used these results when calculating the premiums for stand-alone critical illness insurance charged with only a partial knowledge of family history. They computed adverse selection costs using a Markov model for the insurance market where all lives started uninsured and there could be no lapsing. The transition intensity for purchase after test was allowed to increase for positive results. Similarly, the sum assured for adverse selectors could be at an increased level.

Gui et al. (2006) set up semi-Markov and Markov models for life insurance and critical illness insurance markets respectively. These included the development of family history itself as a state. They calculated the probabilities of developing a family history (two first-degree relatives having BC/OC before age 50) directly for each genotype and from these the transition intensities by summing over the probabilities of each possible event leading to a family history. In their models, genetic testing was only available after the development of a family history. To deal with the semi-Markov nature of life insurance markets, they used the method of Gutiérrez and Macdonald (2004): calculate the expected present values of the reserve at the point of BC/OC development and set this to be the sum assured payable on developing BC or OC. Transition intensities related to BRCA1 and BRCA2 mutations were from models fitted to Antoniou et al. (2003)'s estimates. Mortality after BC/OC diagnosis was independent of the genotype and based on Coleman et al. (1999). Other transition intensities were those used by Macdonald et al. (2003b).

Macdonald and McIvor (2006)'s polygenic model had women inheriting common but small impact genes. Inheritance of these polygenes was not the Mendelian pattern

they used for the major genes; instead parents were each assigned a total number of identical high risk alleles and inheritance was through a hypergeometric distribution based on Lange (1997) (as discussed above). Due to the number of possible genotypes, performing the same summation as Gui et al. (2006) would have taken a vast amount of computer time so they performed simulations to derive their family history details (premium rates and transition intensities). They used the BRCA1 and BRCA2 transition intensities and allele frequencies of Antoniou et al. (2002) since these had been calculated in a model which included the polygenic component. Macdonald and McIvor (2009) applied this model in a market for critical illness insurance, with a state for family history, in a similar manner to Gui et al. (2006) to compare the costs of adverse selection from the polygenes and the major genes.

Viswanathan et al. (2007) allowed lives to react not only to test results, but to the changes in price also. They set out a discrete-time Markov model of annually renewable life insurance based on Subramanian et al. (1999). The quantity of insurance to purchase at the renewal date was such that for price, P, quantity, Q and elasticity of demand, λ, $P^\lambda Q = c$, where c is a constant. Elasticity of demands were estimated by asking a small sample of health care workers how much insurance they would purchase given prices and a higher or lower risk risk of death. Premiums were also allowed to change based on anticipated demand (after taking account of the change in price), such that an equilibrium was reached where the insurer calculated an expected profit of zero.

Lu et al. (2011a) modelled the development of BC in the context of income protection insurance with allowance for only the major genes (BRCA1 and BRCA2). To allow for the possibility of recovery, they included the effects of treatments; while after recovery, they allowed for recurrence. These treatments depended on the type of BC diagnosed so transition intensities to various levels of severity were fitted using medical literature. Their overall BC onset intensities were those fitted by Gui et al. (2006). Lu et al. (2011b) used this model of BC development in order to calculate the associated insurance premiums and policy values. Since their recovery was duration dependent, the model is semi-Markov; in common with Gui et al. (2006), they used Gutiérrez and Macdonald (2004)'s trick of paying the prospective policy value upon transition into an otherwise duration dependent state to bring it back into the Markov framework.

Following on from this work, we will set out a multiple state Markov model for an insurance market. We will draw greatly from Macdonald and McIvor (2006, 2009), the only other studies to include a polygenic component of BC risk.

2.2 The Polygenic Model

In this section we describe the polygenic model of BC risk, which we will use to estimate genotype specific transition intensities for BC onset when we set out our Markov model of female lifetime in Section 2.3. This will further allow us to model adverse selection for lives who have a high (or low) risk of developing BC.

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

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

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

2.2.1 Pharoah's Model

The model of Pharoah et al. (2008) combines multiplicatively the effects of seven known genetic loci to assign a total relative risk of BC. They are assumed to have no effect on OC risk. Since everyone has two functioning alleles of every gene (except those on the X and Y chromosomes), there are three possible combinations of alleles at each locus: no high-risk alleles; one low-risk allele and one high-risk allele; or two high-risk alleles. If the population prevalence of a high-risk allele at locus i is p_i, these three genotypes have population prevalences $(1 - p_i)^2$, $2p_i(1 - p_i)$ and p_i^2 respectively. They report the relative risk of developing breast cancer for each high-risk allele, relative to a low-risk allele, known as the *per allele risk*. Table 2.1 sets out the estimated per allele risk and associated 95% confidence interval of the high-risk variant relative to the low-risk and population prevalences for each of these polygenes with their identification number in the SNP database[1], commonly referred

---

[1]http://www.ncbi.nlm.nih.gov/projects/SNP/

Table 2.1: Per allele risk with 95% confidence intervals and population prevalences of genetic loci known to contribute to the polygenic risk of breast cancer (Pharoah et al., 2008) with the approximate standard deviation of the per allele risk estimate.

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

to as 'dbSNP Number', as well as our own reference label, locus i for $i = 1, 2, \ldots, 7$. For example, an individual with 2 high-risk alleles at both the rs2981582 and the rs1053487 loci has a relative risk of $1.26^2 \times 1.13^2 = 2.0272$ and the probability of this combination is $0.38^2 \times 0.86^2 = 0.1068$. Also given is our calculation of the standard deviation of the estimate, calculated based on a normal approximation since their point estimates for the mean are based on GWAS of large numbers of lives, as

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

(2.1)

2.2.2 Extending The Model

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

polygene.

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

2.2.3 Notation

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

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

2. Choose a starting age x low enough that no cases of BC have occurred. As noted in Antoniou et al. (2001), the proportions of lives in the population with each polygene mutation will change as higher-risk lives get BC and die earlier than lower-risk lives, lowering the average risk of survivors. The population prevalence of major genes and the polygene at age x will be that arising from Mendelian inheritance, with no survivorship effect yet.

3. Let $_tp_x^g$ be the probability that a woman free of BC at age x is still free of it age $x + t$, given she has genotype g.

4. Let p_g be the population prevalence of polygenotype $g \in G$, assumed to be that of our chosen starting age x.

5. Denote the age-dependent population BC transition intensity at age $x + t$ by $\mu^{BC}(x + t)$, and the polygenotype-specific transition intensity at age $x + t$ by $\mu_g^{BC}(x + t)$ for polygenotype $g \in \mathcal{G}$.

6. Define $\lambda(x + t)$ to be the baseline BC transition intensity at age $x + t$. We choose our baseline genotype to be a woman with zero high-risk variants at each locus contributing to the polygene, *i.e.* $\lambda(x + t) = \mu_{(0,0,\ldots,0)}^{BC}(x + t)$.

7. Define RR_{i,g_i} to be the relative risk associated with polygenotype $g_i \in \mathcal{G}_i$ at locus $i \in \{1, \ldots, n\}$, and RR_g to be the relative risk for polygenotype $g \in \mathcal{G}$, assuming that the relative risks associated with genotypes at each contributing genetic locus are constant (so we have proportional hazards). The average relative risk within the population at age $x + t$ is denoted $\overline{RR}(x + t)$.

8. For the ith locus, define

$$\beta_i = \log(RR_{i,1}), \tag{2.2}$$

where $RR_{i,1}$ is the per allele relative risk of the high-risk variant of locus i, given in Table 2.1.

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

$$\mu_g^{BC}(x + t) = \lambda(x + t) RR_g = \lambda(x + t) RR_{(g_1, \ldots, g_n)} = \lambda(x + t) \prod_{i=1}^{n} RR_{i,g_i}. \tag{2.3}$$

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

$$\mu_g^{BC}(x + t) = \mu_{(g_1, \ldots, g_n)}^{BC}(x + t) = \lambda(x + t) \exp(\beta_1 g_1 + \ldots + \beta_n g_n), \tag{2.4}$$

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

2.2.4 Distribution of Relative Risk

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

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

p_i^{*2} respectively. $G_1, G_2, \ldots, G_{147}$ are assumed to be independent random variables and for $j = 1, \ldots, 7$, the random variable G_{j+7k} has the same distribution as G_j, for each $k = 0, 1, \ldots, 20$.

Then the (random) log relative risk associated with locus i can be written as

$$\log RR_{i,G_i} = \beta_i\, G_i. \tag{2.5}$$

We assume that $p_{j+7k}^* = p_j$ for $j = 1, 2, \ldots, 7$ and $k = 0, 1, \ldots, 20$ with the value of p_j given by Table 2.1. For $j = 1, \ldots, 7$, let G_j^* be the random variable representing the total number of high-risk variants in the 21 analogues of locus j, in a woman drawn at random from the population, *i.e.* $G_j^* = \sum_{k=0}^{20} G_{j+7k} \sim \text{Bin}(42, p_j^*)$. This distribution can be well approximated by a normal distribution: $G_j^* \sim \mathcal{N}(42p_j^*, 42p_j^*(1 - p_j^*))$. Moreover, let $RR_{j,G_j^*}^*$ be the random variable representing the total relative risk contributed by the 21 analogues of that locus, in that same woman, *i.e.* $RR_{j,G_j^*}^* = \prod_{k=0}^{20} RR_{j+7k,G_{j+7k}}$ and

$$\log RR_{j,G_j^*}^* = \sum_{k=0}^{20} \log RR_{j+7k,G_{j+7k}} = \sum_{k=0}^{20} \beta_{j+7k}G_{j+7k} = \sum_{k=0}^{20} \beta_j G_{j+7k} = \beta_j G_j^*. \tag{2.6}$$

Now, we have the distribution,

$$\beta_j G_j^* \sim \text{Bin}\left(2 \times 21 \times \beta_j, p_j^*\right). \tag{2.7}$$

Hence, assuming polygenes are inherited and act independently, the total relative risk RR_G of a woman drawn at random from the population has a distribution given by:

$$\log RR_G = \sum_{j=1}^{7} \beta_j G_j^* \sim \mathcal{N}\left(\sum_{j=1}^{7} 42\beta_j p_j^*, \sum_{j=1}^{7} 42\beta_j^2 p_j^*(1 - p_j^*)\right). \tag{2.8}$$

Hence, relative to polygenotype $g = (0, 0, \ldots, 0)$ using the point estimates of relative risk and their associated probabilities from Table 2.1, the polygene relative risk is approximately log-normally distributed with parameters $E(\log RR_G) = 17.4289$ and $\sigma_{\log RR} = 1.20$ at outset, where $\sigma_{\log RR}$ is the standard deviation of $\log RR_G$. This is reasonably close to the simpler model using the hypothetical polygene with three loci (which was used in Macdonald and McIvor, 2006) of Antoniou et al. (2002), who calculated an estimate of 1.29 for the standard deviation with a 95% confidence interval of (1.096, 1.521) using a U.K. population based cohort.

2.2.5 Baseline Rate of Onset of Breast Cancer

In Equation (2.3), $\mu_g^{BC}(x+t)$ is defined in terms of baseline hazard rate, $\lambda(x+t)$. However, $\lambda(x+t)$ is not directly observable. What is observable is the population average onset rate, $\mu^{BC}(x+t)$, which embodies the survivorship effect mentioned in (b) of Section 2.2.3 above. Here we show how to calculate $\lambda(x+t)$ and constrain the average onset rate of BC for survivors at age $x+t$ in the model to be equal to the population onset rate. Noting that the mean relative risk at age $x+t$, $\overline{RR}(x+t)$, is given by:

$$\overline{RR}(x+t) = \frac{\sum\limits_{g\in\mathcal{G}} p_g \, {}_tp_x^g RR_g}{\sum\limits_{g\in\mathcal{G}} p_g \, {}_tp_x^g}, \tag{2.9}$$

we use the relation

$$\lambda(x+t) = \frac{\mu^{BC}(x+t)}{\overline{RR}(x+t)}, \tag{2.10}$$

Our distribution of log relative risk implies a mean relative risk of approximately 7.7×10^7. Hence the baseline onset rate of BC at the starting age x is,

$$\lambda(x+0) \approx \frac{\mu^{BC}(x+0)}{7.7 \times 10^7}. \tag{2.11}$$

To find $\lambda(x+t)$ for $t > 0$, we solve differential equations for ${}_tp_x^g$ for $g \in \mathcal{G}$

$$\begin{aligned}
\frac{d}{dt} {}_tp_x^g &= -\mu_g^{BC}(x+t) \, {}_tp_x^g \\[2mm]
&= -\lambda(x+t) \, RR_g \, {}_tp_x^g \\[2mm]
&= -\frac{\mu^{BC}(x+t)}{\overline{RR}(x+t)} \, {}_tp_x^g \, RR_g \\[2mm]
&= -\frac{\mu^{BC}(x+t) \sum\limits_{h\in\mathcal{G}} p_h \, {}_tp_x^h}{\sum\limits_{h\in\mathcal{G}} p_h \, {}_tp_x^h RR_h} {}_tp_x^g \, RR_g.
\end{aligned} \tag{2.12}$$

Using the approximate distribution derived in Section 2.2.4, the sums over genotypes in Equation (2.12) become integrals, *i.e.*

$$\frac{d}{dt} {}_t\tilde{p}_x^r \approx -\frac{\mu^{BC}(x+t) \int_{-\infty}^{\infty} f_{\log RR_G}(s) \, {}_t\tilde{p}_x^s \, ds}{\int_{-\infty}^{\infty} f_{\log RR_G}(s) \, {}_t\tilde{p}_x^s \exp(s) \, ds} {}_t\tilde{p}_x^r \exp(r). \tag{2.13}$$

where ${}_t\tilde{p}_x^r$ is the probability that a woman free of BC at age x is still free of BC
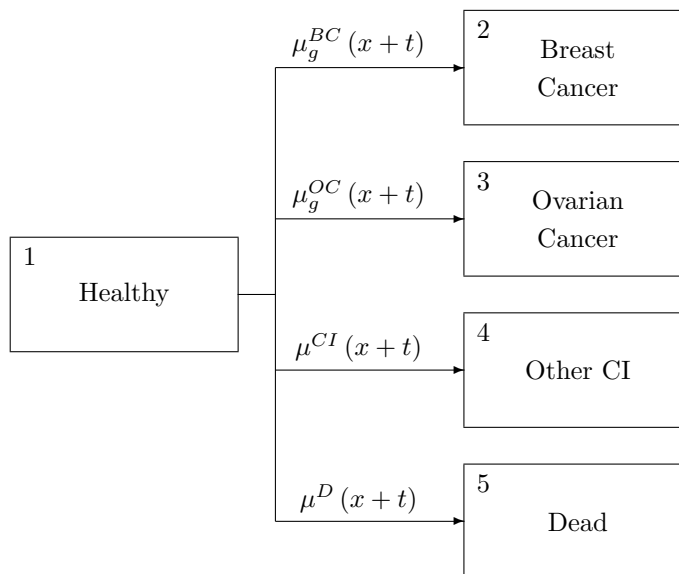
Figure 2.1: A model for critical illness insurance policyholder. Transition to 'Dead' or 'Other CI' states is at rates depending only on age x. The onset rate of OC, $\mu_g^{OC}(x+t)$, depends also on the BRCA genotype but is unaffected by polygenotype part of g, while the onset rate of BC, $\mu_g^{BC}(x+t)$, depends on both the polygenotype and BRCA genotype parts of g.

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

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

2.3 Cost of Family History

2.3.1 Insurance Model

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

The insurance contract of interest is stand-alone critical illness insurance with benefits payable on the occurrence of a defined critical illness, but not on earlier death. This type of product is of interest due to the high proportion of claims that can be directly attributed to BC. It is also the simplest to model as there is no need to consider post-onset survival.

Consistent with previous work (Gui et al., 2006; Macdonald and McIvor, 2006), we define a family history to be one in which two or more first degree relatives have been diagnosed with BC or OC before age 50.

We construct a Markov model for the lifetime of an individual female as illustrated in Figure 2.1. To incorporate the major genes, BRCA1 and BRCA2, we augment the set of possible genotypes to $\mathcal{G}^* = \{0, 1, 2\} \times \mathcal{G}$. Observed overall genotypes are thus $g = (g_0, g_1, \ldots, g_{147})$ where $g_0 \in \{0, 1, 2\}$ is the major genotype corresponding to BRCA0, BRCA1, BRCA2 respectively. For consistency with previous work, transition intensities from 'Healthy' to 'Dead' and 'Other CI' are taken from Gutiérrez and Macdonald (2003). Their mortality rates were based on the English Life Tables No. 15, which were constructed based on the mortality experience of England and Wales over the period 1990–1992. The onset of 'Other CI' was calculated by fitting the major causes of such a claim: stroke, heart attack (both adjusted for lives who do not survive the 28 days required for a claim in a typical CI contract) and cancer (see Appendix B). The transition intensity into the 'Dead' state was adjusted to remove strokes and heart attacks that would result in claim since a claim would have already ceased their policy. The cancer component of 'Other CI' includes the onset of all cancers so onset rates of BC and OC have been removed from them, using transition intensities from Macdonald et al. (2003a) as these were calculated from the same data (Cancer Registrations 1990–1992). BC and OC onset rates are those used in Macdonald and McIvor (2006), calculated from cancer registrations in England and Wales 1983–1987, for consistency with the adjustments for BRCA mutations estimated by Antoniou et al. (2002).

2.3.2   Premiums

Premium rates for each relative risk are calculated as continuously payable level net premiums by solving Thiele's equations using a Runge-Kutta algorithm with step size 2^{-11} years and force of interest 0.05 *per annum*. A summary of these premiums as a percentage of those chargeable to a life with no BRCA mutation and $RR_g = \overline{RR}(x)$ is shown in Table 2.2, allowing us to see how deviations from the population average risk at outset impacts premium rates.

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

$\log RR_g = \log \overline{RR}(x) + 3)$, although the proportion of lives with such a high risk is small.

Table 2.2: Female premium rates as percentage of the premium rates for a life with no BRCA mutation and polygene relative risk $\log RR_g = \log\left[\overline{RR}(x)\right]$.

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

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

The ratings due to polygene vary by age and term, the highest ratings for lives without a BRCA mutation being for 40 year olds with a 10 year term. The reason for this can be seen in Figure 2.2: BC is a significant proportion of the onset rates and this proportion reaches a peak between the ages of 40 and 50 so the polygene has the highest impact on policies with terms that include this age range. The polygene influence decreases as heart attack and other forms of cancer unaffected by the polygene reduce the weighting of BC in the makeup of premium rates. Hence, for lives aged 30 years, there is a peak in ratings for those with a high polygene risk with a 20 year term compared to 10 or 30 year terms. Conversely, since the low polygene risk removes proportionally more of the benefit, there is a dip for lives aged 30 with a low polygene risk for a 20 year term.

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

2.3.3   Simulation of Family Histories

Our aim in this subsection is to model the development of family histories. This will allow us to find the joint distribution of relative risk and major genotype, conditional on family history status and current age. Additionally, we will use the results to calculate family history onset rates in Section 2.5.2.

Recall from Section 2.1.1 that Gui et al. (2006) calculated the probabilities of developing a family history by summation over the possible events. Due to the large number of possible genotypes and inheritance possibilities in our model, it is not tractable to compute onset rates of family history analytically. Instead we simulate the future lifetimes of a large number of lives and their families.

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

1. A family starts with 2 parents, whose polygenotype and BRCA genotype are randomly sampled according to the population distributions in Table 2.1.

Figure 2.2: Stacked onset rates for selected critical illnesses and death of females. Sources: Gutiérrez and Macdonald (2003); Macdonald and McIvor (2006).

Table 2.3: Probability mass function of number of daughters in a family conditional on there being at least one daughter. Source: Macdonald et al. (2003a).

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

2. The number of daughters is randomly sampled from a random variable with the probability mass function given in Table 2.3. We use the distribution estimated by Macdonald et al. (2003a). This was based on Shaw (1990)'s estimate of the number of children born to women who were themselves born in England and Wales in 1940–44, extrapolated according to the proportions of children born to women who were married in 1961–65 according to the 1971 census.

3. Daughters are the same age and born when the mother, assumed to be healthy, is age 30.

4. Each daughter inherits polygenes and BRCA genes from their parents independently of each other according to Mendel's laws, acting at each major gene locus and each polygene locus. The validity of this assumption could be questioned — it is conceivable that some of the polygene loci might be within close proximity to another on the same chromosome and exhibit some level of linkage.

5. Per-allele relative risk for locus i, for $i = 1, 2, \ldots, 147$, is sampled from a normal distribution with standard deviation σ_i and mean equal to the point estimate listed in Table 2.1 and held constant for all lives in the simulation. In this way we include the sampling variance from Pharoah et al. (2008) in our modelling, as well as the sample variance from our simulation.

6. Full life histories of mother and daughters are simulated using their respective genotypes and treating each of the decrements in Figure 2.1 as independent competing risks.

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

The simulation of a large number of lives (within 15,000,000 families) allows us to observe the distribution of relative risk at any age x, and how it depends on family history.

2.3.4 Distribution of Relative Risk by Underwriting Class

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

For each life, at every age $x+t$, we know her genotype, whether she is healthy and which relatives, if any, have so far been diagnosed with BC or OC before age 50. We categorise this large quantity of data at each age by: BRCA genotype; $\log RR_g$; and presence of family history. Since $\log RR_g$ is a continuous quantity we discretise it into 100 'bins' of uniform length 0.125 over the range $[\log \overline{RR}(20) - 6.25, \log \overline{RR}(20) + 6.25]$. We choose this range because it is large enough that the probability of a life having a relative risk outside it, is 1.924×10^{-7}. For family history categorisation, we define two underwriting classes

1. ST — Unrated — Females without a family history.

2. FH — Rated — Females who have developed a family history.

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

We ran the model 500 times to produce confidence intervals for the estimates of the means and standard deviations of $\log RR_G$. The resulting discretised distributions for the average of our simulations are shown in Figure 2.3 for ST and FH at ages $x = 30$ and 50 for each genotype. (Note that Figure 2.3 includes male lives which we need in simulations in Section 2.4.)

The distribution of relative risk of BRCA0 lives within each underwriting class, remains roughly the same at each age despite higher risk lives leaving the class (by either getting cancer or developing a family history). There is very little difference in the spread of $\log RR_G$ in ST and in FH; $\sigma^{FH}_{\log RR}(x)/\sigma^{ST}_{\log RR}(x)$ is close to 1 (see Table 2.5). However, as can be seen from the plots and Table 2.4, there is a significant difference in the mean of $\log RR_G$ in each class, with $E^{FH}_{\log RR}(x) - E^{ST}_{\log RR}(x)$ between 1.1 and 1.2. This results in average relative risk within FH being approximately 250% higher than within ST at ages 30, 40 and 50, making it a good proxy for genetic risk. To check convergence, we compare to estimates based on 100 simulations, shown in Tables 2.6 and 2.7. As these show little difference at ages of interest, we continue with the 500 simulations we have performed.

For lives with a BRCA mutation, the differences between the mean of $\log RR_G$ in each underwriting class are lower than that of BRCA0 and decreases with age. While

Table 2.4: Mean of log relative risk for female lives in underwriting classes ST and FH based on 500 simulations.

BRCA	Age (x)	$E^{ST}_{\log RR}(x)$	95% Confidence Interval	$E^{FH}_{\log RR}(x) - E^{ST}_{\log RR}(x)$	95% Confidence Interval
0	20	17.43	$(17.24, 17.64)$	0.85	$(0.40, 1.35)$
	30	17.43	$(17.24, 17.64)$	1.18	$(1.05, 1.33)$
	40	17.43	$(17.23, 17.63)$	1.24	$(1.20, 1.29)$
	50	17.41	$(17.22, 17.61)$	1.16	$(1.13, 1.19)$
1	20	17.43	$(17.24, 17.65)$	0.46	$(-1.29, 2.98)$
	30	17.43	$(17.24, 17.65)$	0.89	$(-0.29, 2.23)$
	40	17.32	$(17.14, 17.51)$	0.86	$(0.73, 1.03)$
	50	17.18	$(17.01, 17.36)$	0.56	$(0.48, 0.65)$
2	20	17.43	$(17.23, 17.65)$	0.81	$(-1.00, 2.97)$
	30	17.43	$(17.23, 17.65)$	0.99	$(-0.23, 2.39)$
	40	17.34	$(17.16, 17.54)$	1.01	$(0.87, 1.18)$
	50	17.22	$(17.04, 17.39)$	0.82	$(0.74, 0.93)$

there are insufficient lives with a family history at age 30 to draw solid conclusion, there are significant differences between FH and ST at ages 40 and 50. The spread of $\log RR_G$ in FH is slightly smaller than ST for BRCA2 at age 50. However at all other ages, and for BRCA1, there are no significant differences.

The proportions of lives with a BRCA mutation in each of the underwriting classes is also significantly different, as can be seen in Table 2.8.

2.3.5 Notation

We introduce notation to be used in this section. Suppose a life aged x has applied for a policy with term n years.

1. $f_{\log RR, G_0 | x, FH}(r, g)$ is the joint probability density function for $\log RR_G$ and major genotype, G_0, conditional on a life aged x having developed a family history of BC or OC. For succinctness, this is abbreviated to $f^{FH}_x(r, g)$. A similar expression is used for ST.

2. $A_{r,g,x:\overline{n|}}$ is the expected present value of 1 paid on transition to a state other than Healthy or Dead for a life with relative risk e^r and major genotype g currently aged x.

3. $a_{r,g,x:\overline{n|}}$ is the expected present value of an annuity of 1 paid continuously while in the Healthy state for a life with relative risk e^r and major genotype g currently aged x.
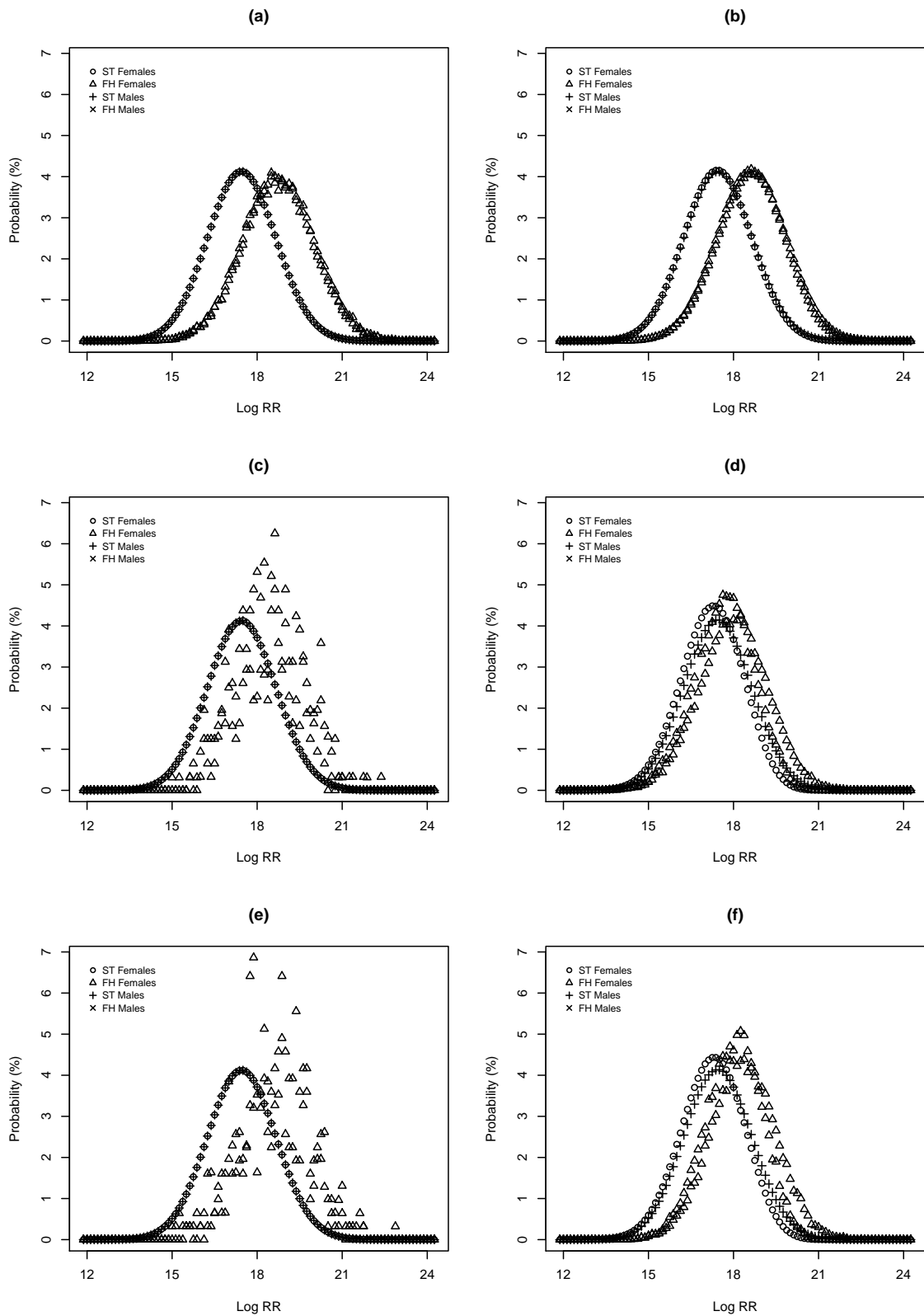
Figure 2.3: Distributions of relative risk for females and males: (a) BRCA0, age 30; (b) BRCA0, age 50; (c) BRCA1, age 30; (d) BRCA1, age 50; (e) BRCA2, age 30; and (f) BRCA2, age 50, averaged over 500 simulations.

Table 2.5: Standard deviation of log relative risk for female lives in underwriting classes ST and FH based on 500 simulations.

BRCA	Age (x)	$\sigma^{ST}_{\log RR}(x)$	$\sigma^{FH}_{\log RR}(x)/\sigma^{ST}_{\log RR}(x)$	95% Confidence Interval
0	20	1.203	1.009	$(0.767, 1.292)$
	30	1.202	1.045	$(0.963, 1.135)$
	40	1.199	1.016	$(0.991, 1.046)$
	50	1.188	1.000	$(0.984, 1.017)$
1	20	1.203	0.336	$(0.065, 1.258)$
	30	1.202	0.677	$(0.197, 1.411)$
	40	1.141	0.946	$(0.867, 1.030)$
	50	1.083	0.976	$(0.910, 1.026)$
2	20	1.203	0.250	$(0.062, 1.478)$
	30	1.202	0.441	$(0.114, 1.472)$
	40	1.154	0.928	$(0.841, 1.027)$
	50	1.097	0.905	$(0.843, 0.972)$

Table 2.6: Mean of log relative risk for female lives in underwriting classes ST and FH based on 100 simulations.

BRCA	Age (x)	$E^{ST}_{\log RR}(x)$	95% Confidence Interval	$E^{FH}_{\log RR}(x) - E^{ST}_{\log RR}(x)$	95% Confidence Interval
0	20	17.43	$(17.25, 17.60)$	0.874	$(0.459, 1.423)$
	30	17.43	$(17.25, 17.59)$	1.179	$(1.064, 1.347)$
	40	17.42	$(17.25, 17.59)$	1.196	$(1.157, 1.243)$
	50	17.40	$(17.23, 17.57)$	1.112	$(1.084, 1.140)$
1	20	17.43	$(17.25, 17.60)$	0.742	$(-0.810, 2.676)$
	30	17.43	$(17.25, 17.60)$	0.898	$(0.000, 2.003)$
	40	17.31	$(17.14, 17.46)$	0.874	$(0.750, 1.040)$
	50	17.17	$(17.02, 17.31)$	0.555	$(0.467, 0.651)$
2	20	17.43	$(17.26, 17.60)$	0.560	$(-0.241, 2.213)$
	30	17.43	$(17.26, 17.60)$	1.065	$(0.001, 2.344)$
	40	17.34	$(17.18, 17.49)$	1.010	$(0.876, 1.190)$
	50	17.20	$(17.06, 17.35)$	0.820	$(0.729, 0.950)$

Table 2.7: Standard deviation of log relative risk for female lives in underwriting classes ST and FH based on 100 simulations.

BRCA	Age (x)	$\sigma^{ST}_{\log RR}(x)$	$\sigma^{FH}_{\log RR}(x)/\sigma^{ST}_{\log RR}(x)$	95% Confidence Interval
0	20	1.203	1.025	$(0.792, 1.307)$
	30	1.203	1.047	$(0.966, 1.130)$
	40	1.199	1.011	$(0.989, 1.011)$
	50	1.188	1.004	$(0.991, 1.019)$
1	20	1.203	0.336	$(0.133, 1.085)$
	30	1.203	0.677	$(0.242, 1.448)$
	40	1.142	0.946	$(0.871, 1.023)$
	50	1.085	0.976	$(0.926, 1.032)$
2	20	1.202	0.245	$(0.086, 0.712)$
	30	1.202	0.425	$(0.175, 1.376)$
	40	1.154	0.930	$(0.830, 1.028)$
	50	1.098	0.905	$(0.847, 0.977)$

Table 2.8: Proportions of lives with a BRCA mutation.

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

Table 2.9: Distribution of premium rates charged to lives with a family history as a percentage of the premium charged to lives without a family history.

Age	Term	Mean (%)	95% Confidence Interval (%)	Standard Deviation
	10	192.7	$(162.4, 225.2)$	0.1661
30	20	187.3	$(162.7, 212.6)$	0.1262
	30	168.5	$(150.4, 186.9)$	0.0930
40	10	247.3	$(235.4, 259.4)$	0.0596
	20	202.0	$(193.2, 210.3)$	0.0419
50	10	158.6	$(154.8, 162.5)$	0.0204

4. $\Pi^{ST}_{x:\overline{n}|}$ is the level premium for a life without a family history.

5. $\Pi^{FH}_{x:\overline{n}|}$ is the level premium for a life with a family history.

2.3.6 Ratings for Presence of Family History

With these distributions of polygene relative risk and of major gene mutations, we will calculate the level premiums that would cover the cost of CI benefits payable assuming underwriters may use family history as a factor.

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

$$\Pi^{FH}_{x:\overline{n}|} = \frac{\sum_{g=0}^{2}\left[\int_{-\infty}^{\infty} f_x^{FH}(r,g) A_{r,g,x:\overline{n}|}\,dr\right]}{\sum_{g=0}^{2}\left[\int_{-\infty}^{\infty} f_x^{FH}(r,g) a_{r,g,x:\overline{n}|}\,dr\right]} \tag{2.14}$$

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

We use the estimated distributions based on each of our 500 simulations to estimate premium rates for lives with a family history and show how uncertainty may cause them to vary. Results of this are shown in Table 2.9 and Figure 2.4. They show a narrow confidence interval for age 50 at entry. However younger ages at entry have a broader interval, up to 65% of the unrated premium.

The approach taken by Macdonald and McIvor (2006) was to compare weighted premiums for lives with a family history (as calculated above) to the premiums appropriate for a life with BRCA0 and polygene relative risk $RR_g = 1$. However, the average polygene risk within the unrated pool will be somewhat higher — they used a binomial distribution for the log risk, centred at 0 — which increases the average premium for lives without a family history, offsetting some of the extra payable for the lives with a family history. In a reconciliation exercise, this was found to be around
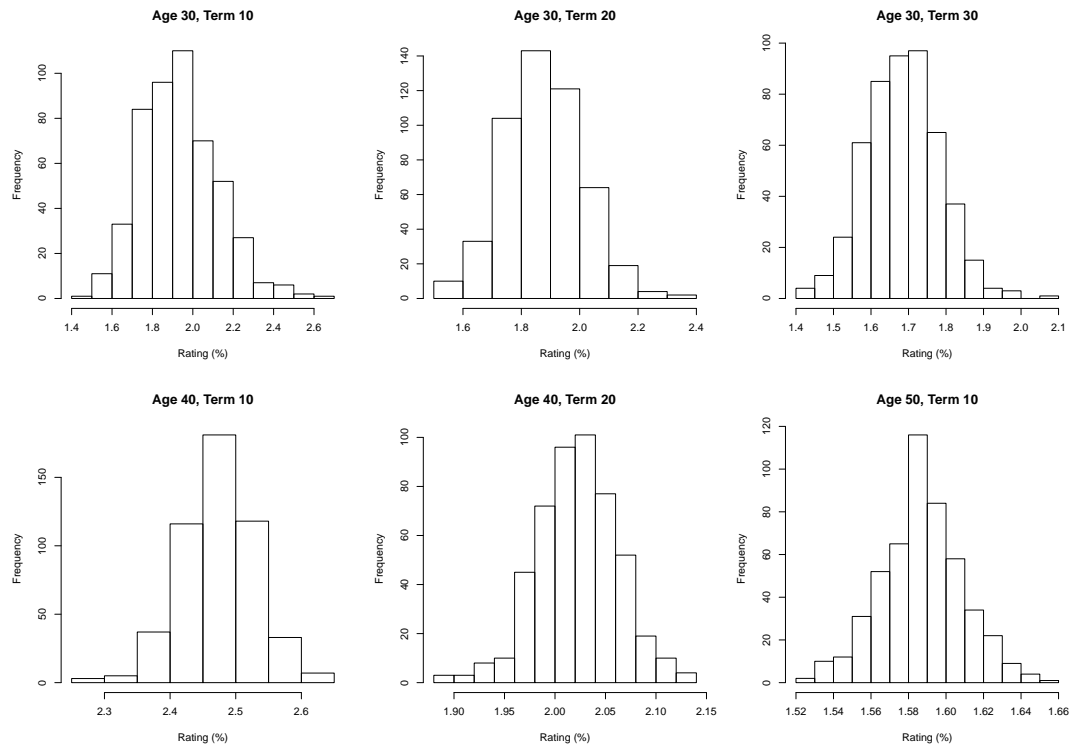
Figure 2.4: Histograms of ratings for presence of family history as a percentage of the premium charged to lives without a family history.

40%.

Allowing for this change, and comparing our average ratings in Table 2.10, the results from the two polygene models do show some differences at ages at entry 40 and 50 of 30–40% while age 30 is far greater at all terms by 100–250%. Macdonald and McIvor ran their model only once and at age 30 they had 91 lives with a family history, 9 of which had a BRCA mutation as well as polygene relative risk greater than 1, resulting in large premium ratings. It is therefore possible that the difference between our results and theirs at age 30 is down to sampling. In either case, the largest part of rating is attributable to the polygenes.

2.4 Effect of the 'Test-Achats' European Court of Justice Ruling

Recall from Chapter 1 that the European Court of Justice ruled that pricing with sex as a rating factor would not be permitted from 22nd December 2012. We assume that insurers' response will be to charge premiums based on aggregate risk on an assumed business mix. It is not clear whether this extends to underwriting based on justifiable biological risk and it is possible that this would need to be tested in court.

Table 2.10: Level net premiums for females with a family history of BC or OC as a percentage of those for females without a family history, averaged over simulations. Also shown are equivalent results from Macdonald and McIvor (2006). Note MG results use P+MG for experience basis but MG only for pricing basis.

| Study | Genetic Model | Age 30 | | | Age 40 | | Age 50 |
		10 yrs %	20 yrs %	30 yrs %	10 yrs %	20 yrs %	10 yrs %
Ours	P+MG	193	187	169	247	202	159
	MG	110	108	106	149	133	113
M&M (2006)	P+MG	444	341	275	244	207	171
	MG	138	132	123	113	109	103

Table 2.11: Probability mass function of the number of children in a family. The probability that a particular child is born female is 1/2.06. Source Macdonald et al. (2003a).

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

For example it is not clear whether or not a man, who has a family history of BC among his female relatives should be given the same rating his sister would be given. There are two possibilities:

1. Females receive a rating reflecting the average risk of a female with a family history and male lives are not rated.

2. Males and females are rated equally for a family history of BC or OC based on the aggregate risk of their sales mix.

The model described above needs to be adjusted to include male lives. Each family is randomly assigned a number of children following the distribution of family size and sex calculated by Macdonald et al. (2003a) (see Table 2.11).

Research by Tai et al. (2007) shows an increased risk of male BC due to BRCA mutations, particularly BRCA2. However, detailed research is limited, partly due to the rarity of BC in males. BRCA mutations have also been linked to prostate cancer

(see Narod et al., 2008) but this is not usually a severe form of cancer and diagnosis is often missed until later in life. Consequently it would not be particularly helpful in identifying potential BRCA carriers from family history. We therefore perform simulations and calculate male premiums ignoring BRCA mutation in males except for transmission from father to daughter. Transition rates for males are those given in Gutiérrez and Macdonald (2003) (see Appendix B).

2.4.1 Notation

To distinguish between males and females we introduce further notation

1. $f_x(r, g)$ is the joint distribution function for $\log RR_G$ and and major genotype, G_0.

2. $f_{FH|F}(x)$ and $f_{FH|M}(x)$ are the probabilities that a female and male respectively, develop a family history of BC or OC by age x.

3. $A^F_{x:\overline{n}|}$ and $A^M_{x:\overline{n}|}$ are the expected present values of 1 paid to a life aged x in underwriting class ST immediately on transition to a claim state for females and males respectively for a term of n years.

4. $a^F_{x:\overline{n}|}$ and $a^M_{x:\overline{n}|}$ are the expected present values of a continuously payable annuity at rate 1 *per annum* while in an insured state, to a life currently aged x and in underwriting class ST for females and males respectively for a term of n years.

5. $A^F_{r,g,x:\overline{n}|}$ and $a^F_{r,g,x:\overline{n}|}$ are the female log relative risk and major genotype specific assurance and annuity functions defined above in Section 2.3.6.

6. $\Pi'^{FH}_{x:\overline{n}|}$ is the premium when males with a family history are rated with females.

7. We denote the probabilities a newborn is female or male as $p_f = 1/2.06$ and $p_m = 1.06/2.06$ respectively.

8. $P(M)$ and $P(F)$ are the proportion of males and females respectively who buy our insurance contract.

2.4.2 Preliminary Analysis

Table 2.12 gives the premium rates as a percentage of those for a female with $\log RR_g = \log \overline{RR}(x)$ and no BRCA mutation.

The percentages of females and males who develop a family history in our simulations are shown in Table 2.13. At each age there is no significant difference between the percentages of females with a family history of BC or OC and those of males.

Table 2.12: Premium rates for males as a percentage of premium rates for females with $\log RR_g = \log RR_x$ and no BRCA mutation.

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

Table 2.13: Percentage of healthy lives with a family history of breast or ovarian cancer shown by sex and age.

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

For calculation of unisex premiums we set an even split between the sexes at each age of our insured population: $P(M) = P(F) = 0.5$. In reality the respective weightings of each sex will depend on the target market of the insurer.

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

$$A^F_{x:\overline{n}|} = \sum_{g=0}^{2} \left[\int_{-\infty}^{\infty} f_x^{ST}(r, g) \, A^F_{r,g,x:\overline{n}|} dr \right],\tag{2.15}$$

and the expected present value of her continuously payable annuity as

$$a^F_{x:\overline{n}|} = \sum_{g=0}^{2} \left[\int_{-\infty}^{\infty} f_x^{ST}(r, g) \, a^F_{r,g,x:\overline{n}|} dr \right],\tag{2.16}$$

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

$$\Pi^{ST}_{x:\overline{n}|} = \frac{A^F_{x:\overline{n}|} P(F) + A^M_{x:\overline{n}|} P(M)}{a^F_{x:\overline{n}|} P(F) + a^M_{x:\overline{n}|} P(M)}.\tag{2.17}$$

Note that setting the proportion of males, $P(M) = 0$ results in the premiums calculated above in Equation (2.14).

Table 2.14: Distribution of premium rates charged to female lives with a family history as a percentage of the unisex premium charged to lives without a family history.

Age	Term	Mean (%)	95% Confidence Interval (%)	Standard Deviation
	10	207.6	$(172.0, 247.3)$	0.1983
30	20	193.4	$(166.6, 222.0)$	0.1415
	30	167.4	$(149.2, 186.3)$	0.0961
40	10	250.4	$(236.2, 265.7)$	0.0741
	20	196.0	$(186.8, 206.5)$	0.0484
50	10	151.5	$(147.3, 156.3)$	0.0233

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

2.4.3 Females receive a rating, Males do not

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

$$
\Pi^{FH}_{x:\overline{n}|} = \frac{
P(F) \sum_{g=0}^{2} \int_{-\infty}^{\infty} f_x(r,g) \left(A^F_{r,g,x:\overline{n}|} - \Pi^{ST}_{x:\overline{n}|} a^F_{r,g,x:\overline{n}|} \right) dr + P(M) \left(A^M_{x:\overline{n}|} - \Pi^{ST}_{x:\overline{n}|} a^M_{x:\overline{n}|} \right)
}{
f_{FH|F}(x) P(F) \sum_{g=0}^{2} \int_{-\infty}^{\infty} f^{FH}_x(r,g) a^F_{r,g,x:\overline{n}|} dr
}
$$

$$
+ \ \Pi^{ST}_{x:\overline{n}|}. \tag{2.18}
$$

In comparing Tables 2.9 and 2.14, the effect of unisex premiums is to increase ratings slightly at younger ages where male CI costs are lower than females — reducing the premiums that females pay — and to decrease ratings slightly at older ages where males have high heart attack and stroke rates (see Figure 2.5) — increasing the premiums that females pay.
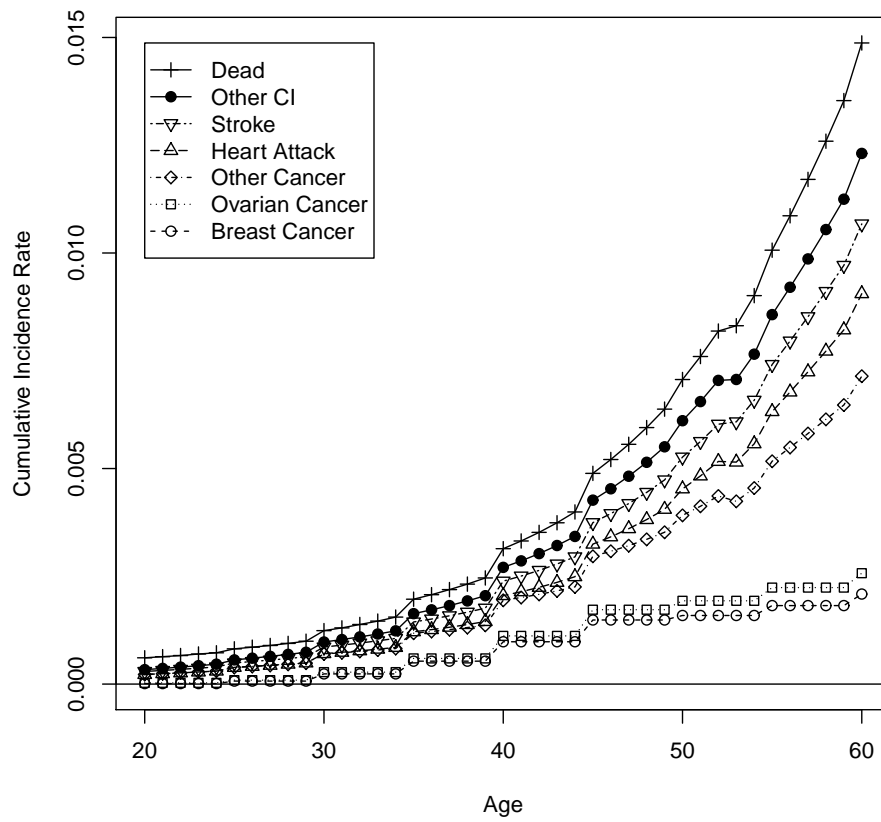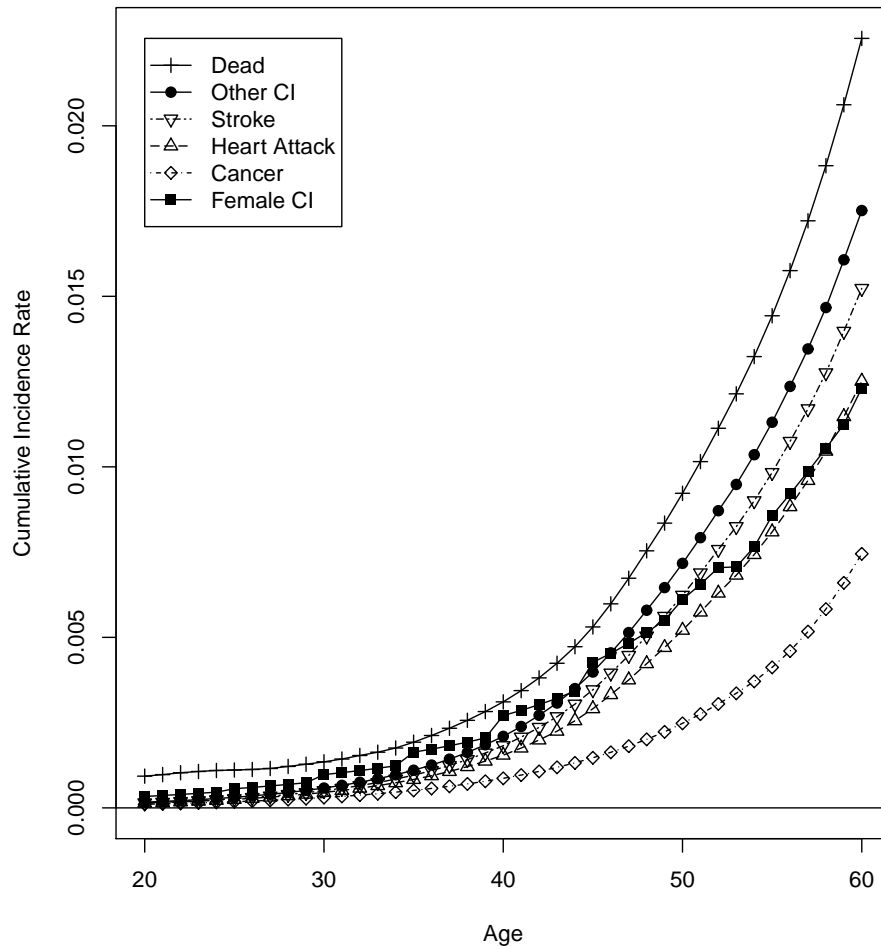
Figure 2.5: Stacked onset rates for selected critical illnesses and death in males. Source: Gutiérrez and Macdonald (2003). The population average female CI rate from Figure 2.2 is overlaid for comparison.

Table 2.15: Distribution of premium rates charged to male and female lives with a family history as a percentage of the unisex premium charged to lives without a family history.

Age	Term	Mean (%)	95% Confidence Interval (%)	Standard Deviation
	10	153.8	$(136.4, 174.0)$	0.1010
30	20	146.5	$(133.4, 159.9)$	0.0719
	30	133.4	$(124.3, 142.2)$	0.0487
40	10	173.0	$(165.9, 179.3)$	0.0355
	20	146.3	$(141.8, 150.7)$	0.0227
50	10	124.8	$(122.9, 126.9)$	0.0106

2.4.4 Females and Males receive a rating

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

$$
\Pi'^{FH}_{x:\overline{n}|} = \frac{P(F) \sum_{g=0}^{2} \int_{-\infty}^{\infty} f_x(r,g) \left(A^F_{r,g,x:\overline{n}|} - \Pi^{ST}_{x:\overline{n}|} a^F_{r,g,x:\overline{n}|} \right) dr + P(M) \left(A^M_{x:\overline{n}|} - \Pi^M_{x:\overline{n}|} a^M_{x:\overline{n}|} \right)}{f_{FH|F}(x) P(F) \sum_{g=0}^{2} \int_{-\infty}^{\infty} f^{FH}_x(r,g) a^F_{r,g,x:\overline{n}|} dr + f_{FH|M}(x) P(M) a^M_{x:\overline{n}|}}
$$

$$
+ \quad \Pi^{ST}_{x:\overline{n}|}. \tag{2.19}
$$

Since the extra cost of a high proportion of high-risk female lives in the FH underwriting class is spread over approximately twice the number of lives than when only females are rated, it is not surprising that the ratings and standard deviations in Table 2.15 are approximately half those in Table 2.14.

2.5 Calculating Adverse Selection In An Insurance Market

With the identification of these seven loci, the possibility of testing for an individual's overall BC risk is one step closer. If a life has a pre-symptomatic genetic test, then under the UK's moratorium on the use of genetic information in underwriting, the

life is not required to divulge these test results. This asymmetry of information presents the life with an opportunity to change their buying behaviour. Lives who see themselves as low-risk may buy less insurance, while lives who see themselves as high-risk may buy more insurance. Both are forms of adverse selection.

Recall from Section 1.2 that researchers have found adverse selection in insurance markets where individuals have private information: Brown (1992) in health insurance; Finkelstein and Poterba (2004) in annuities markets; and Finkelstein and McGarry (2006) in a market for long-term care. Armstrong et al. (2003) found carriers of BRCA1 or BRCA2 were significantly more likely to increase their life cover. However, Zick et al. (2000) suggest that women who have tested positive for BRCA1 mutations are not more likely to purchase more life cover than untested women. Neither are women who have tested negative for BRCA1 less likely to purchase life cover than untested women. They suggest that their follow-up period is too short since many participants were still dealing with the health impacts of receiving a positive test result.

A common outcome of positive BRCA mutation tests is the removal of breasts (mastectomy) and ovaries (salpingo-oophorectomy) as a means to reduce risk. Skytte et al. (2010) found over a 10-year period, 50% of mutation carriers had a preventative mastectomy and 75% salpingo-oophorectomy. However, polygenes have not yet got such a clinical use so a woman who knows she is high risk might not have prophylactic surgery and instead purchase CI.

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

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

2.5.1 Market Model

We expand our insurance model to an insurance market, incorporating buying behaviour that may change due to the result of genetic testing. The product of interest is again stand-alone critical illness. To ensure the model is Markov, the premium is payable continuously as a current risk premium instead of a level premium. This risk premium at time t is calculated for each sex as the weighted average cost of expected claims arising in $(t, t + dt)$, with weights equal to occupancy probabilities in insured states assuming no adverse selection — all lives purchase insurance at the same standard rate. To conform to the EU legislation, these gender-specific risk premiums will have to be averaged across the two sexes with the weights for each being the overall probability a male and female is insured respectively, calculated assuming adverse selection to take place.
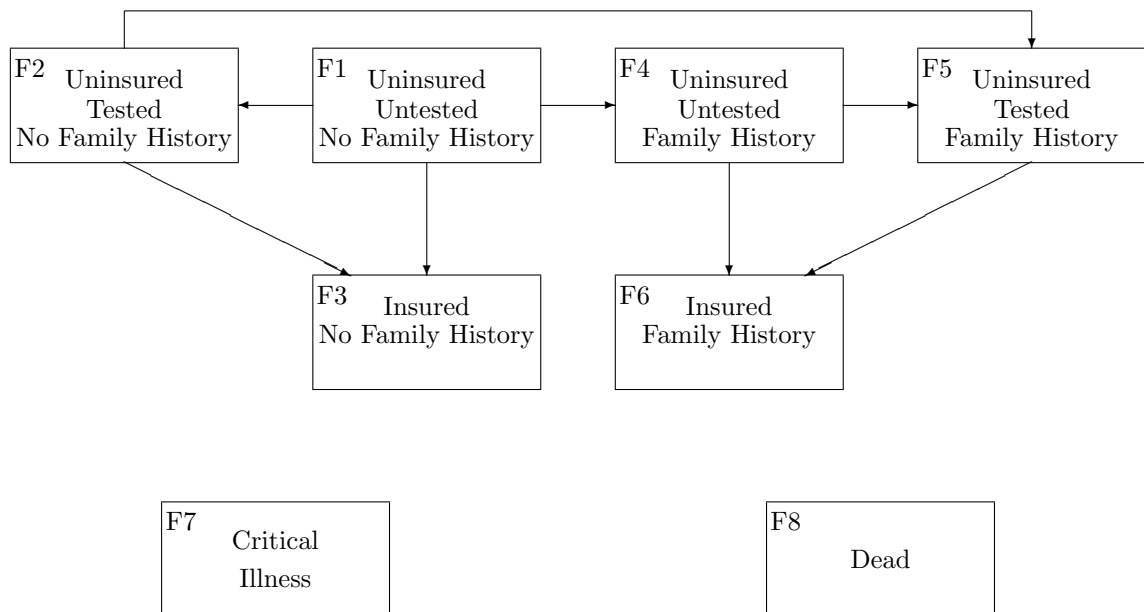
Figure 2.6: A model of female insurance states for an insurance market where genetic testing may be available before and after family history at different rates. The arrows to Dead and Critical Illness states are omitted but these may be entered from any other state.

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

All lives are assumed to be 20 years old, healthy, untested and uninsured at the start of our modelling.

The model is illustrated in Figures 2.6 and 2.7.

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

1. The rates of going into a tested state are those used by Macdonald and McIvor (2009) which they based on Ropka et al. (2006)'s estimate of 59% uptake: High with rate 0.08916 *per annum*; Medium with rate 0.04458 *per annum*; or Low with rate 0.02972 *per annum*. All testing is performed on lives between ages 20 and 40.

2. Market size is at two levels represented by the rate at which a standard life enters the insured state: a Large market with rate 0.05 *per annum* or Small market with 0.01 *per annum*.

3. After receiving test results, a life may change their buying behaviour. Low risk lives either buy at the same standard rate, half the standard rate or do not buy
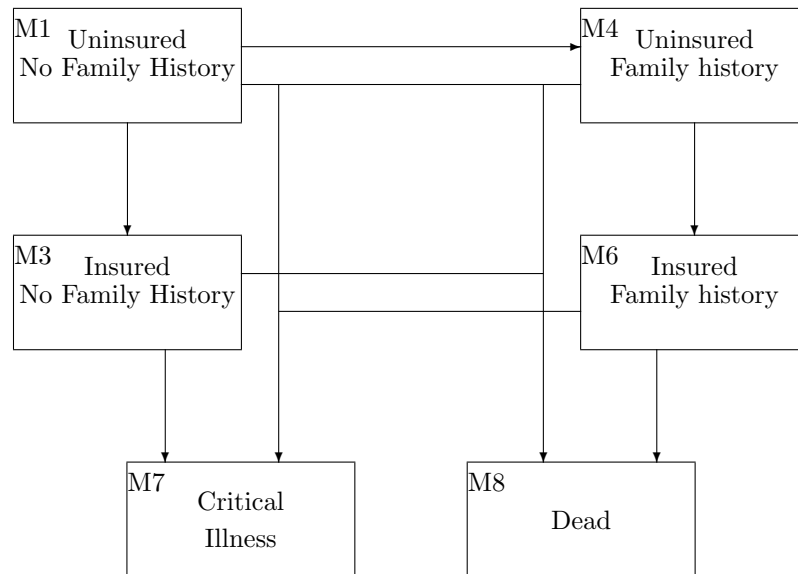
Figure 2.7: A model of male insurance states in an insurance market. Depending on legislation, lives in state state M6 will belong to either underwriting class ST or FH.

at all. High risk lives, regardless of the size of market, will buy at rate 0.25 *per annum* which we take to represent a 'Severe' level of adverse selection; or at rate 0.1 *per annum* which we consider to be 'Moderate' adverse selection. A life is deemed to be high risk if their log relative risk is higher than the log of the population average relative risk plus the threshold in the scenario (see below). They are low risk if their risk is lower than the log of the population average relative risk minus the threshold. Additionally, a life with a mutation in either BRCA1 or BRCA2 is always deemed high risk, regardless of the polygene relative risk.

The threshold of relative risk serves to classify lives into three categories of relative risk. It is varied to show the effect of how different from the population risk a life needs to see herself as being before she considers herself to be high risk or low risk *i.e.* whether the product represents high or low value to her. We limit ourselves to using high-risk and low-risk thresholds that are symmetric around the log of population average relative risk. However, in reality, the distance from this average that is needed to change behaviour may differ for low-risk and high-risk lives. We choose average relative risk as a central point because the premium charged will be approximately equal to the appropriate premium for a BRCA0 life with average relative risk. When tested lives consider whether they are getting a cheap or expensive rate, it will be with reference to this average premium. However, this will create an imbalance between the proportions of high and low risk lives (see Table 2.16). Macdonald and McIvor (2009) used the mode of their distribution as the centre.

This could be enhanced with the use of some decision making methodology *e.g.*

Table 2.16: Proportion of lives at age 20 considered Low or High Risk at each threshold point.

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

utility theory or prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) to assess whether the product represents good value. Previous work has aimed to find the potential for adverse selection using utility theory by considering boundaries for which lives would not buy insurance; they do not consider how lives would buy at an increased rate because their expected utility framework only answers the question whether or not to buy. For instance Macdonald and Tapadar (2010) modelled a multi-factorial gene disorder that interacts with the environment and found no realistic circumstances where healthy lives would choose not to buy insurance. Whereas Macdonald and McIvor (2009) found that under two of the four utility models used (the same as those parameterised in Macdonald and Tapadar (2010)), lives with a very low risk polygenotype would not buy insurance.

2.5.2   Family History Onset Rate

The market model includes the onset of family history as an event. We define the onset rate of family history as

$$\text{Onset Rate} = \frac{\text{New family history cases in period}}{\text{Exposed to risk}}, \tag{2.20}$$

where the exposed to risk is the sum over all lives of the number of years the life is healthy and has not yet developed a family history in the period. As can be seen in Figures 2.8, 2.9 and 2.10, the log of the onset rate has a strong linear dependence on the individual's log polygene risk. However, there is divergence from the straight line at each end of the plots, particularly in the BRCA1 and BRCA2 families. This is due to the very small number of lives with the extremes of relative risk. With a small value for the denominator in Equation (2.20) for each additional family history there is a large jump in the calculated onset rate. We therefore omit groups where there are less than 10 lives contributing to the exposure. There is a large difference where a parent carries a BRCA mutation so these are considered separately. As there is no observable difference between sexes, we ignore sex as a factor. Family history cannot arise after the life is age 50 as the definition is limited to cancer before age 50 and all lives are assumed to be born at the same time.
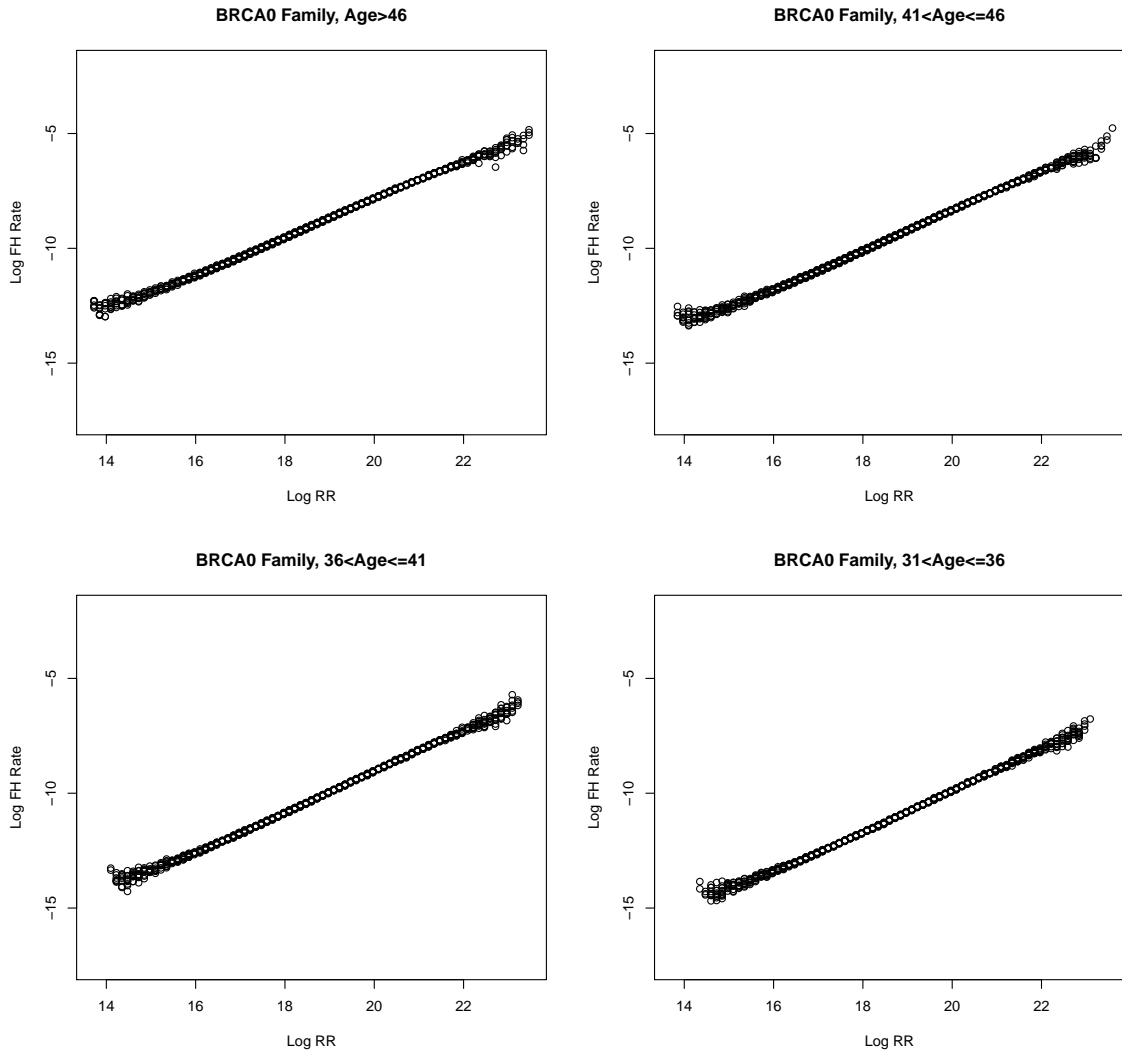
Figure 2.8: Logarithm of family history onset rates for families with no BRCA mutation in either parent. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.

Based on the linearity of the log of the onset rate, we fit linear models for the log of the onset rate as the response variable, using simulations from Section 2.4, with log relative risk and age as explanatory variables and Normal errors. Since the sizes of the simulated onset rates are small, this is unlikely to be associated with high adverse selection costs, we will therefore be satisfied with simple predictor functions. We choose a predictor with quadratic terms for age and log relative risk and interactions between each, and choose variables to be included in an iterative process by analysing the fit and removing the least significant term, subject to the deviance not increasing a significant amount, until we are left with p-values$< 5\%$ for each term (Crawley, 2012).

Denote the BRCA0 specific onset rate of family history as $\mu_{\text{BRCA0}}^{H}(r, x)$ for a life
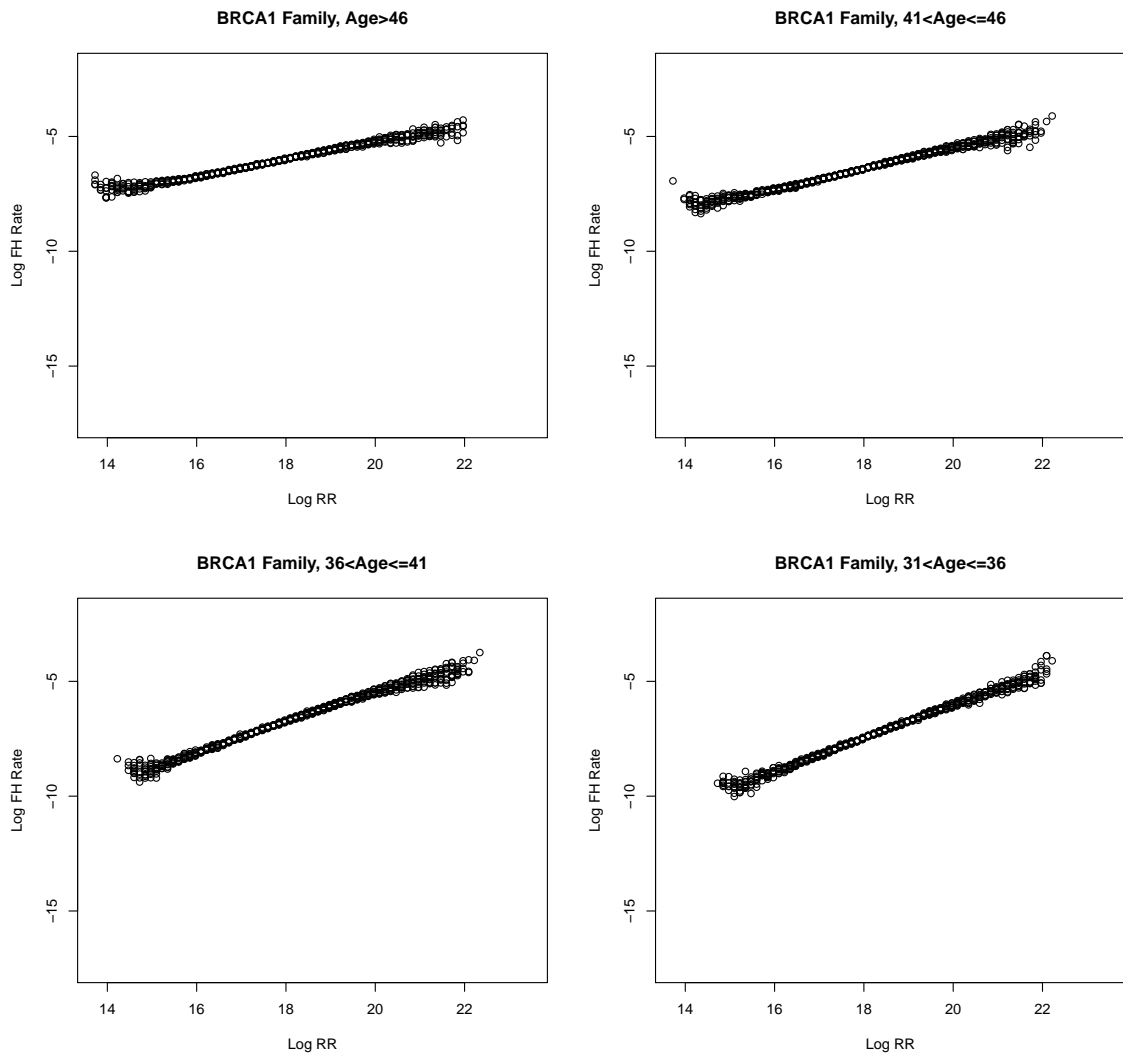
Figure 2.9: Logarithm of family history onset rates for families where a parent has a BRCA1 mutation. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.
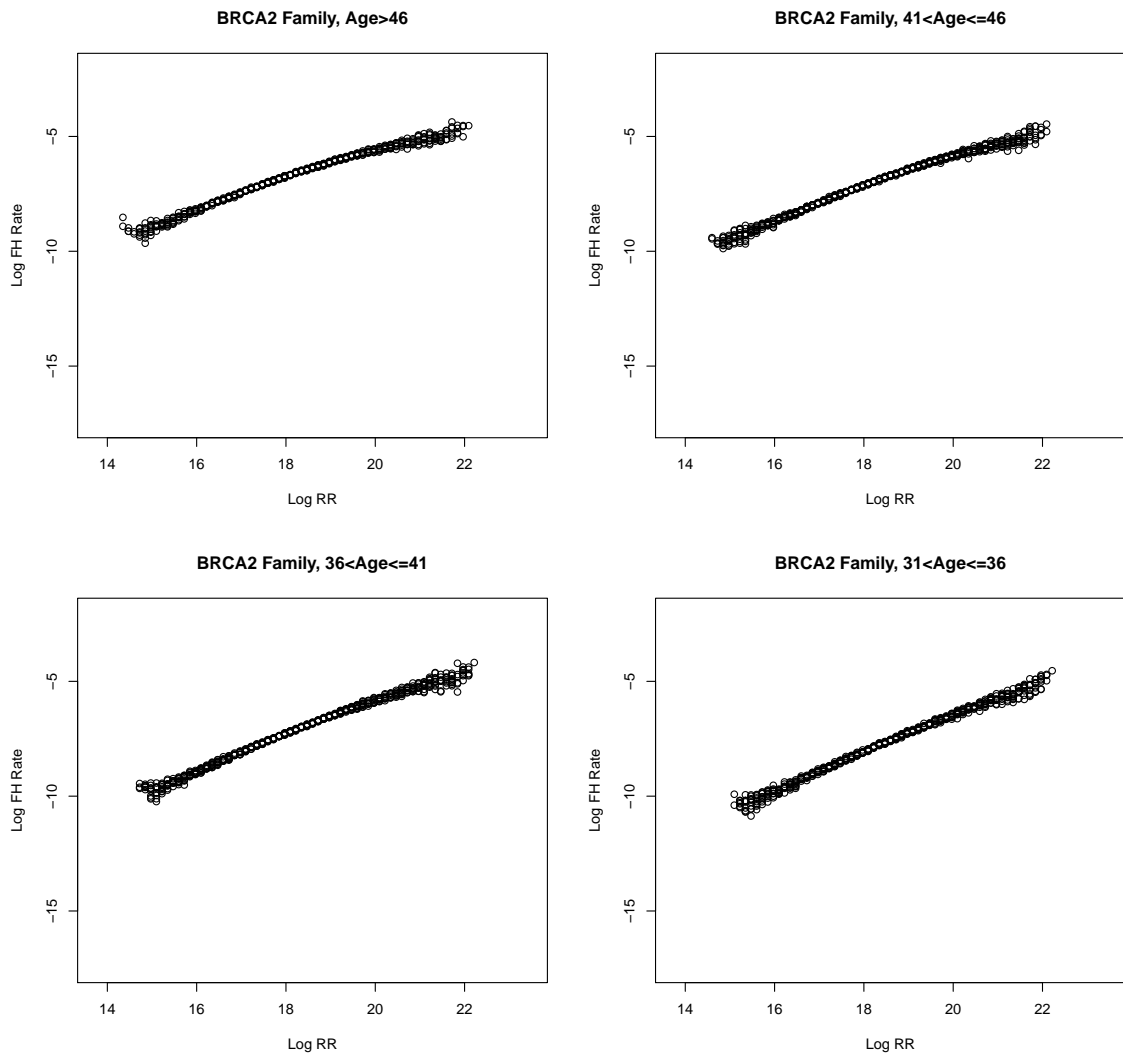
Figure 2.10: Logarithm of family history onset rates for families where a parent has a BRCA2 mutation. Each circle represents the calculated onset rate for a particular age, sex and relative risk combination. Groups of size less than 10 are omitted from the plots.

aged x with $\log RR_g = r$ and similarly for BRCA1 and BRCA2.

For BRCA0 this results in the following,

$$
\mu^H_{\text{BRCA0}}(r, x) = \begin{cases} \exp(-21 - 0.4696x + 0.0081x^2 + 0.0463xr - 0.0006x^2r), & \text{if } x < 50 \\ 0, & \text{if } x \geq 50. \end{cases}
$$
(2.21)

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

$$
\mu^H_{\text{BRCA1}}(r, x) = \begin{cases} \exp(-12.44512 - 4.44898x + 0.2486xr), & \text{if } x \leq 30 \\ \exp(-37.63246 + 4.34126x - 0.2344xr + 1.5182r), & \text{if } 30 < x < 50 \\ 0, & \text{if } x \geq 50 \end{cases}
$$
(2.22)

and

$$
\mu^H_{\text{BRCA2}}(r, x) = \begin{cases} \exp(-12.44512 - 4.44898x + 0.2486xr), & \text{if } x \leq 30 \\ \exp(-31.44989 + 0.20735x - 0.011xr + 1.17576r), & \text{if } 30 < x < 50 \\ 0, & \text{if } x \geq 50. \end{cases}
$$
(2.23)

2.5.3 Cost of adverse selection

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

$$
\frac{\text{E(PV Benefits|Adverse Selection)} - \text{E(PV Premium Income|Adverse Selection)}}{\text{E(PV Premium Income|Adverse Selection)}},
$$
(2.24)

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

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

$$
{}^g_t p^{j,k}_x = P(\text{In State } k \text{ at age } x + t | G^* = g \text{ and in State } j \text{ at age } x).
$$
(2.25)

They may be calculated by solving the Kolmogorov forward equations (see Appendix

A):

$$\frac{d}{dt}{}^{g}_{t}p^{j,k}_{20} = \sum_{l \neq k} \left({}^{g}_{t}p^{j,l}_{20}\mu^{l,k}_{20+t} - {}^{g}_{t}p^{j,k}_{20}\mu^{k,l}_{20+t} \right). \tag{2.26}$$

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

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

The continuously payable risk premium rate at time t may then be found using a similar technique to that used to calculate the premiums in Equation (2.17), by calculating the mean critical illness hazard rate for female lives in ST, $\overline{\alpha}^{F3}_{x+t}$, and including lives in FH when averaging across the sexes:

$$\Pi^{ST}(t) = \frac{\sum\limits_{g \in \mathcal{G}^*} p_g \left[p_f \overline{\alpha}^{F3}_{20+t} \left({}^{g}_{t}p^{F1,F3}_{20} + {}^{g}_{t}p^{F1,F6}_{20} \right) + p_m \overline{\alpha}^{M3}_{x+t} \left({}^{g}_{t}p^{M1,M3}_{20} + {}^{g}_{t}p^{M1,M6}_{20} \right) \right]}{\sum\limits_{g \in \mathcal{G}^*} p_g \left[p_f \left({}^{g}_{t}p^{F1,F3}_{20} + {}^{g}_{t}p^{F1,F6}_{20} \right) + p_m \left({}^{g}_{t}p^{M1,M3}_{20} + {}^{g}_{t}p^{M1,M6}_{20} \right) \right]}. \tag{2.27}$$

Lives in states $M3$ and $F3$ (see Figures 2.7 and 2.6) belong to the same underwriting class, ST, due to the requirement of unisex rates. However, the question over how to treat males with a family history among their female relatives is still present. Again there are two possible cases:

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

$$\Pi^{FH}(t) = \overline{\alpha}^{F6}_{20+t} - \overline{\alpha}^{F3}_{20+t} + \Pi^{ST}(t); \tag{2.28}$$

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

$$\Pi'^{FH}(t) = \frac{\sum\limits_{g \in \mathcal{G}^*} {}^{g}_{t}p^{F1,F6}_{20} p_g p_f \left(\overline{\alpha}^{F6}_{20+t} - \overline{\alpha}^{F3}_{20+t} \right)}{\sum\limits_{g \in \mathcal{G}^*} p_g \left({}^{g}_{t}p^{F1,F6}_{20} p_f + {}^{g}_{t}p^{M1,M6}_{20} p_m \right)} + \Pi^{ST}(t). \tag{2.29}$$

While these two possibilities may present difficulty to an insurer's underwriting department, the proportion of lives that develop a family history represents only 0.09% of male lives at its peak at age 50 from Table 2.13, so when testing is available to all lives, the impact of underwriting is negligible and we give results for only females

being underwritten.

The expected present value of benefits and premiums needed for Equation (2.24) are then calculated with occupancy probabilities found in the presence of adverse selection, by numerical integration using Simpson's formula as follows:

$$
\begin{aligned}
\text{E(PV Benefit|Adv. Sel.)} \;=\; & \sum_{g \in \mathcal{G}^*} p_g \left[p_f \int_0^{40} e^{-\delta t} \left({}^g_t p_{20}^{F1,ST} + {}^g_t p_{20}^{F1,FH} \right) \alpha_{g,20+t}^F \, dt \right. \\
& \left. + p_m \int_0^{40} \left({}^g_t p_{20}^{M1,ST} + {}^g_t p_{20}^{M1,FH} \right) \alpha_{g,20+t}^M dt \right]
\end{aligned}
\tag{2.30}
$$

and

$$
\begin{aligned}
\text{E(PV Premium|Adv. Sel.)} \;=\; & \sum_{g \in \mathcal{G}^*} p_g \left[p_f \int_0^{40} e^{-\delta t} \left({}^g_t p_{20}^{F1,ST} + {}^g_t p_{20}^{M1,ST} \right) \Pi^{ST}(t) \, dt \right. \\
& \left. + p_m \int_0^{40} e^{-\delta t} \left({}^g_t p_{20}^{F1,FH} + {}^g_t p_{20}^{M1,FH} \right) \Pi^{FH}(t) \, dt \right],
\end{aligned}
\tag{2.31}
$$

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

2.5.4   Results

We choose scenarios which allow us to show the impact of successively adding sources of adverse selection to analyse where the adverse selection costs can be attributed.

The likely testing arrangements are by no means certain. Testing for BRCA1 or BRCA2 mutations from the National Health Service is available only after it is apparent there is risk of the woman having a mutation and always requires some family history[2]. However, since inheritance of polygene risk does not necessarily result in the pattern of inheritance where an individual can be assigned a reasonable probability of having a high risk genotype, a similar testing regime may not be as effective. As the science develops and more genes are identified, the feasibility and cost of testing will improve. Pharoah et al. (2008) discusses the use of population testing as a way of identifying high-risk women in order to target the more expensive means of detecting breast cancers, *e.g.* MRI screening, for early diagnosis. They found that the use of a few susceptibility alleles could be used to distinguish woman as high risk could improve the efficiency of BC screening programmes.

This is being further investigated in a trial across Greater Manchester by the Pre-

---

[2] `http://www.nhs.uk/Livewell/Breastcancer/Pages/Breastcancergenes.aspx` 4th August, 2013

dicting Risk Of Cancer At Screening study (PROCAS) (Howell et al., 2012). PROCAS is using a combination of family history, lifestyle, breast density and genetics to estimate breast cancer risk. Women who are deemed to be high risk are then given advice on how to reduce their risk and offered more frequent monitoring.

If the testing is extended in this way across the country, the potential for adverse selection increases greatly, with most lives being aware of their risk profile. How well this information would be interpreted by the average person is debatable — without genetic counselling, the implications of the results would be unlikely to be properly understood: high risk lives becoming fearful which could impair the rational decision making process (Loewenstein, 2000). Early detection will however change the shape of rate tables and these early cases could be at a stage that is not covered under the critical illness cover *e.g.* ductal carcinoma in situ is not considered a critical illness by many insurers, but left untreated may develop into invasive breast cancer. Where the prognosis improves such that a claim would not be viable, the inclination to adversely select would decrease so we include the Moderate adverse selection rate to reflect this case.

We include scenarios representing an extension of the current National Health Service policy to include polygene testing after a family history has been observed, and where testing is available to all women regardless of family history.

The results of these scenarios are shown in Tables 2.17 to 2.21:

1. High risk lives adversely select at a Severe rate after a polygene test (Table 2.17).

2. Low risk lives change behaviour after a polygene test, while high lives risk buy at the standard rate (Table 2.18).

3. High risk lives adversely select at a Severe rate after tests for polygene and BRCA — lives with BRCA1 or BRCA2 mutations are considered high risk, regardless of polygene relative risk (Table 2.19).

4. High risk lives adversely select at a Moderate rate after tests for polygene and BRCA — lives with BRCA1 or BRCA2 mutations are considered high risk, regardless of polygene relative risk (Table 2.20).

5. High risk lives adversely select at a Severe rate after tests for polygene and BRCA — lives with BRCA1 or BRCA2 mutations are considered high risk, regardless of polygene relative risk. Tests available only to lives with a family history of breast or ovarian cancer (Table 2.21).

Table 2.17: Cost of severe adverse selection. Testing for PG only is available to all lives, regardless of the presence of a family history of BC or OC. Buying behaviour changes when the log of an individuals relative risk is outside of the range $\log\left[\overline{RR}(x)\right] \pm$ Threshold. Underwriting is performed on females only.

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

Table 2.18: Cost of adverse selection when high-risk lives buy at the standard rate. Testing for PG only is available to all lives, regardless of the presence of a family history of BC or OC. Buying behaviour changes when the log of an individuals relative risk is outside of the range $\log\left[\overline{RR}(x)\right] \pm$ Threshold. Underwriting is performed on females only.

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

Table 2.19: Cost of severe adverse selection. Testing for MG and PG is available to all lives, regardless of presence of a family history of BC or OC. Buying behaviour related to polygenes changes when the log of an individuals relative risk is outside of the range $\log\left[\overline{RR}\left(x\right)\right] \pm$ Threshold. Underwriting is performed on females only.

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

Table 2.20: Cost of moderate adverse selection. Testing for MG and PG is available to all lives, regardless of presence of family history of BC or OC. Buying behaviour related to polygenes changes when the log of an individuals relative risk is outside of the range $\log\left[\overline{RR}(x)\right] \pm$ Threshold. Underwriting is performed on females only.

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

Table 2.21: Cost of severe adverse selection. Testing for MG and PG is available only after the development of a family history of BC or OC. Buying behaviour related to polygenes changes when the log of an individuals relative risk is outside of the range $\log\left[\overline{RR}\left(x\right)\right] \pm$ Threshold. Underwriting is performed on females only.

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

The costs involved look comparable to those in Macdonald and McIvor (2009). However our model differs from theirs in a number of ways, key of which is the European Union ruling described in Section 2.4. This spreads the costs over approximately twice the volume of premium. Thus the costs suggested by our model would need to be approximately doubled to apply to a regime where sex discrimination is permitted. Where this is the case, our somewhat moderate costs could become quite substantial particularly when passed on to the customer.

One observation of note is that a large part of the cost of adverse selection could come from low risk lives buying at a lower rate after testing. Although the reduction in the portfolio of insured lives by any individual low risk life increases the average risk by a smaller amount than the gain of a high risk life, the volume of low risk lives underbuying compared to high risk lives overbuying causes this high cost relating to low risk lives. This is somewhat in contrast to Subramanian et al. (1999) who found that the lapsing of women who test negative for a BRCA mutation was of a much lower impact than women with a positive test result increasing benefit levels for a life insurance contract. Their work did not include the polygenic model of BC risk, which introduces wide heterogeneity within the major gene groupings.

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

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

In a small market, under the Moderate adverse selection scenarios there are still reasonable costs involved, up to 15%. Indeed the change from the Severe rate of insurance purchase is quite small as this is already a very high rate of insurance purchase compared to the Standard purchase rate.

2.6   Conclusions

The discovery of a small number of 'breast cancer polygenes' has allowed us to take another look at Macdonald and McIvor (2006)'s work which previously used a hypothetical model in order to assess the findings, and update their results as necessary. They had suggested the simplifications of that model create a heavy tail that distorts the costs. We have found that their results for family history ratings, when corrected to use the standard premium as numeraire, are reasonable. Additionally their conclusion that the bigger part of adverse selection risk is attributable to polygenes still stands in a model based on actual genetic data. However they understate the size of the adverse selection cost compared to that arising under our actual gene model.

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

Chapter 3

Modelling Long-Term Care

3.1 Introduction

In this chapter we set out models to allow us to assess the costs of adverse selection in a start-up market for long term care insurance (LTC) and assess the relative impact of different sources of adverse selection in this market. LTC provides the policyholder with a benefit to cover the cost of payments to a carer or care home. In Chapter 4 we will apply this modelling to a U.K. setting.

Previous actuarial models of genetic adverse selection (including our work in Chapter 2) have assumed an established insurance market. In such markets, premiums have already absorbed the cost from an increased proportion of high-risk lives, and we merely find by what factor they are larger than if there was no adverse selection. Since we are considering a start-up market, we will assume that insurers initially have insufficient experience to set premiums appropriately. As they gain experience, they will compare actual cashflows with what was expected, and respond to differences by changing premiums chargeable to new business. In this way, we make the premium rates dynamic. As the premiums adjust, so will the costs of adverse selection and our model will be able to chart these costs over time. This methodology, where we explicitly model the emergence of information and its relation to adverse selection costs, could also be applied to an established market where a new source of adverse selection is being introduced.

There are two main reasons for requiring care: reduced functional ability and reduced cognitive function (dementia). We will start by describing these causes, then review previous models of long-term care.

LTC is a product to pay for old-age care, so our modelling will commence with detailing a model for old-age health in Section 3.2. This will involve estimating transition intensities between states in a Markov model. We will also introduce and parameterise an intermediate state for dementia, where the life has noticed the initial signs of the disease but a lack of clinical diagnosis creates an information asymmetry

and hence an opportunity for adverse selection.

Under a LTC contract, benefit payments may depend on path taken by the insured through the health model. In order to approximately calculate the expected present values of future benefits at the start of each calendar year of a policy's existence, in Section 3.3 we specify a model to simulate future lifetimes and the resulting LTC cashflows.

In Section 3.4 we will set out a model for a start-up insurance market. In this market model, we will describe how we will use the expected present values calculated from our simulations in order to calculate our dynamic premium rates.

3.1.1 Functional Ability

We measure reduction in functional ability in terms of activities of daily living (ADLs) *e.g.* the Barthel and Katz indices (Mahoney and Barthel, 1965; Katz et al., 1970), and instrumental activities of daily living (IADLs) — these are not as fundamental as ADLs but still necessary to retain one's independence — that can no longer be performed. The ADLs used to trigger an insurance claim are typically needing assistance with at least two of bathing, dressing, toileting, transferring, continence, or feeding. We use the same language as is used among insurers to describe the level of disability:

- A person that has problems performing x ADLs is said to have xADLs;

- A person that can perform all ADLs is said to have no ADLs.

Once functional ability has been lost, it is not necessarily permanent and lives can recover to healthier states. In this way it is similar to income protection insurance and much like income protection, a desirable feature from an insurer's perspective may be a deferred period to prevent the high cost of claims underwriting when the claimant does not require a long period of care.

3.1.2 Cognitive Function

Dementia is an umbrella term for a group of disorders which impair cognitive function by affecting thought processes, memory, judgement and personality. Although the main sufferers are the elderly, some diseases occur at younger ages. It is progressive, so symptoms gradually worsen but sufferers may show these symptoms inconsistently: they may have periods of lucidity despite severe progress of the disease. As it progresses and the sufferer loses the ability to take care of him/herself, informal care from family members or formal care in a nursing home may be necessary.

Diagnosis is aided by use of a mini-mental state examination (MMSE) (Folstein et al., 1975) — a series of 30 questions with particular focus on orientation to time

and place, memory and arithmetic — where lower scores indicate a decreased level of cognitive ability. However, this test cannot easily distinguish between cause of dementia (Santacruz and Swagerty, 2001).

We class the causes of reduced cognitive function into two groups: dementia due to Alzheimer's Disease (AD) and dementias that are not caused by Alzheimer's Disease (Non-AD).

Alzheimer's Disease

The main type of dementia is AD which is responsible for 65–70% of cases (Berr et al., 2005). There is a great deal of overlap in the symptoms shown with those of non-AD, so diagnosis of AD can only be made with certainty at autopsy where the brain may be properly analysed. Characteristics within the brain include a build-up of proteins on the outside of neurones (β-amyloid plaques), inside of neurones (hyperphosphorylated tau protein) and a loss of neurones and synapses.

Its causes are not known but it is thought to have a strong genetic component; Gatz et al. (2006) estimates that between 60–80% of the risk is genetic.

There are two distinct varieties of AD, named for the timing at which they occur: early-onset Alzheimer's disease, which usually affects lives aged less than 60 years old and late-onset Alzheimer's disease, the more common form and which affects lives aged 60 years and older. Early-onset AD should not be confused as merely being the earlier occurrence of AD as the known causes are different — specifically, the genes known to cause early-onset AD do not have an impact on late-onset AD (Gerrish et al., 2012).

Early-onset AD is commonly referred to as Mendelian AD because the pattern of inheritance it generally follows is Mendelian. The genetics in this variety is well established (albeit not completely) with strong evidence supporting the main causes of early-onset AD as being mutations of the genes which code for the amyloid precursor protein, presenilin 1 and presenilin 2 (Alagiakrishnan et al., 2012). These gene mutations are rare but result in a very high probability of developing AD. They cause high production rates of amyloid beta peptides which clump together into the amyloid plaques of Alzheimer's disease.

However, the genetics of late-onset AD is not so well understood and the only gene known to affect its development and that has sufficient epidemiology for modelling purposes is the apolipoprotein E gene, APOE. In the advent of genome-wide association studies, identification of potential susceptibility genes has picked up pace and there have been many candidates for association with late-onset AD. The progress of the research is reviewed in Alagiakrishnan et al. (2012): susceptibility genes for which results have been replicated, giving strong evidence for their association with late-onset AD, are the phosphatidylinositol-binding clathrin assembly protein (PICALM),

clusterin (CLU), complement receptor 1 (CR1), B1N1 and GRB2-associated-binding protein 2 (GAB2) genes. Coon et al. (2007) established APOE as "the major susceptibility gene" for late-onset AD.

There is also a potential role of epigenetics in the risk of developing late-onset AD. Histone modification is a process which can cause genes to either be expressed or silenced and "abnormal methylation pathways are detected in the brains of people afflicted with dementia" (Alagiakrishnan et al., 2012). Environmental factors can also contribute to epigenetic gene expression for example there is suggestion of interaction between APOEε4 variant and cholesterol in Chandra and Pandav (1998). Epigenetics could therefore play a role in altering the risk associated with genetic risk. However, we do not include this in our modelling as the data required to fit relevant parameters is unavailable.

Since we are concerned with post-retirement care, we are only interested in late-onset AD. Therefore, when we refer to AD, this is specifically the late-onset variety.

Non-Alzheimer's Dementia

The various diseases that cause dementia, other than AD, include but are not limited to

- Vascular dementia (also known as multi-infarct dementia) — usually caused by small strokes and is the second most common cause of dementia (Battistin and Cagnin, 2010);

- Dementia with Lewy bodies — Lewy bodies are aggregates of protein in nerve cells;

- Parkinson's disease — usually presents with motor symptoms, caused by degeneration of the central nervous system, before dementia develops later;

- Huntington's disease — a genetic disorder, however symptoms usually show much earlier than the post-retirement ages we are concerned with so this would not add to the adverse selection costs in the type of insurance product we are concerned with.

3.1.3   Previous Models of Long-Term Care

Here we review the methodology and the aims of some of the models of long-term care in the U.K. The major difference between the methodologies is with regards to the use of either transition intensities or prevalence rates.

The first studies we consider used a transition intensity based approach. Transition intensities are a more fundamental quantity than prevalence rates — with transition

intensities into and out of a state, one can calculate prevalence at any given time. The benefit of using transition intensities is that they allow greater flexibility in the modelling — the prevalence of diseases can change overtime and using transition intensities enables this to happen. However they are more difficult to come by — whereas prevalence requires only a snapshot of a population at one point, to estimate transition intensities, we need to understand how the population changes over time. Ideally, such an exercise would involve revisiting the population at a later time to find the details of any changes in health, increasing the cost and duration of analysis.

Macdonald and Pritchard (2000) set out a Markov model for the onset of AD dependent on the APOE variants carried by the life. They had a state for where the AD had progressed to the stage of requiring care in an institution (residential care home), but data available to them was limited for their transition intensities post AD diagnosis. They use this model in Macdonald and Pritchard (2001) in the context of an established market for LTC to calculate potential adverse selection costs from high-risk genotypes buying at an increased rate. As a proxy for the onset of a claim, they consider the trigger for a claim to be the transition into an institution. They calculated single premiums for each genotype using Norberg (1995)'s equations and averaged these across genotype with weights equal to the proportion of lives assumed to buy insurance. Adverse selection cost was calculated as the percentage increase in premiums after increasing the proportion of lives of high-risk genotypes among those who buy the contract.

Pritchard (2006) fitted a Markov model for disability with 5 levels of functional disability to the results of the National Long Term Care Study in the U.S (Manton, 1988). His aim was to estimate the costs of disability claims in a LTC contract. He calculated the expected present value of the benefits attributable to occupancy of each state and found that where studies exclude recovery, they could substantially overstate the cost of benefits.

Akodu (2007)'s Markov model of functional ability and cognitive function was based on the Cognitive Function and Ageing Studies I (CFAS) data. CFAS collected their data by conducting interviews with over 13,000 participants in a longitudinal study MRC CFAS (1998). Akodu (2007) did not consider dementia types separately, but instead classed any life with MMSE score below a certain point as cognitively impaired. To perform a sensitivity analysis on how dementia was defined, models were fitted for MMSE scores below 10, 18, 20 and 21 which were chosen to conform to classifications given by McNamee (2004), Neale et al. (2001) and Spiers et al. (2005) respectively. It was not an insurance model therefore its aim was not to measure adverse selection; instead it was estimating the future demand for long-term care in the U.K. based on projected population sizes of different states.

We now move onto studies which used a prevalence data based approach. A

drawback of using such data is that it requires an assumption of static prevalence, whereas in reality the pattern may change over time.

Similarly, the Personal Social Service Research Unit (PSSRU) of the London School of Economics and the University of Kent, have performed various modelling exercises with regards to long-term care demand and expenditure using updated versions of Wittenberg et al. (1998)'s spreadsheet based model *e.g.* Hancock et al. (2007); Wittenberg et al. (2006). They split the population by risk-factors: age, gender, dependency, household type, housing tenure (as a proxy for economic circumstances) and whether in receipt of informal care. The number of males and females in each age band were projected forward using the U.K.'s Office for National Statistics' and Government Actuary's Department's projections. These were further split into cells for the remaining factors and a probability of receiving formal care attached by fitting functions to the General Household Survey results. A development to Wittenberg et al. (1998) of note was by Comas-Herrera et al. (2003). This was by amending the probability of receiving formal care to include services for cognitive impairment specifically.

Nuttall et al. (1994) used Office of Population Censuses and Surveys prevalence data to fit a discrete-time multiple state model of disability in the U.K.. The aim was to project future demand and costs of care and assess the implications on different sectors' ability to finance LTC. They used 3 states to represent the health of lives, with separate models for different levels of disability. To fit their transition intensities they assumed prevalence of disability was unchanging, although as they point out, their calculated intensities contradict this assumption.

Rickayzen and Walsh (2002) extended Nuttall et al. (1994)'s model to allow transition between ADL states, including recovery. Annual probabilities of transition between states and trends for how these might change over time, were derived from the General Household Survey, Government Actuary's Department projections and Office of Population Censuses and Surveys data. They used this to project the number of disabled people with no regard to care costs.

We would like to develop a multiple state model, based on elements of Macdonald and Pritchard (2001)'s and Akodu (2007)'s models, which uses transition intensities and includes both cognitive and functional disabilities.

3.2   Model of Old Age

In this section we set out our model of old age which will form the basis of our LTC models in Sections 3.3 and 3.4. We start with a continuous-time Markov model of old age as illustrated in Figure 3.1 and parameterise the transition intensities between states.
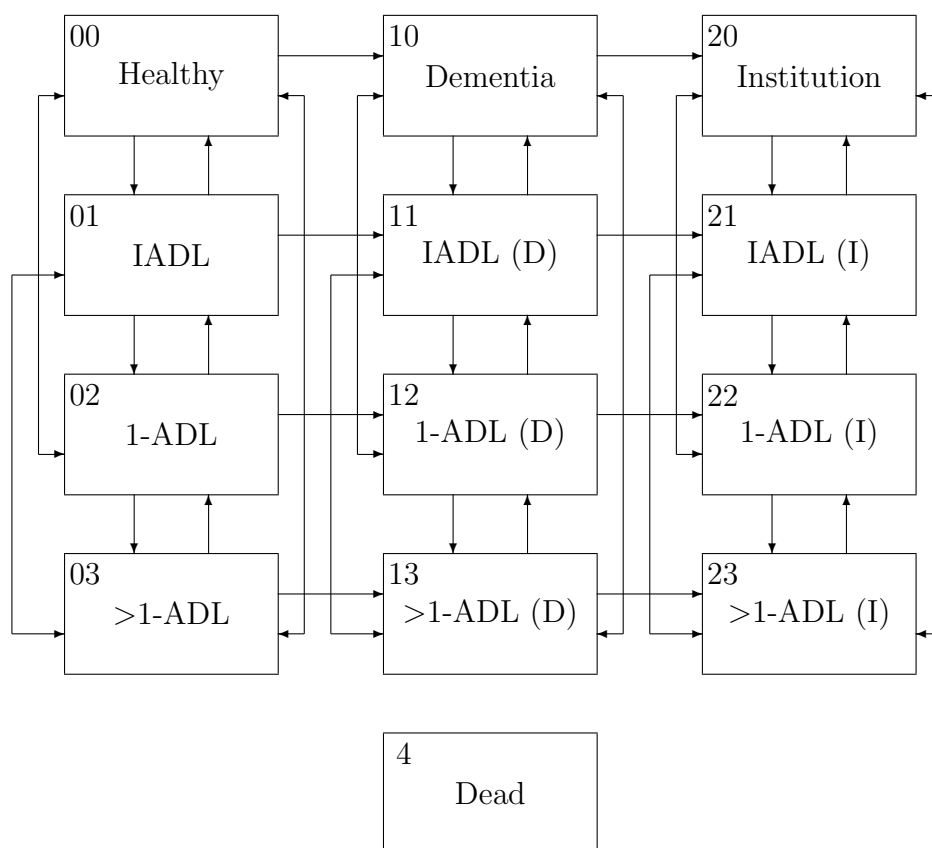
Figure 3.1: A Markov model of functional ability and cognitive function. The arrows to the Dead state are omitted but may be entered from any state.

The Markov framework is an essential assumption for our modelling methodology and is very common in actuarial modelling of life histories. Intuitively, the transitions might be expected to contradict such an assumption. However, there is a difficulty in fitting anything more complex as the data required, such as duration of state occupancy and the path taken to the current state, are often unavailable. Macdonald and Pritchard (2000) discusses the Markov nature of the mortality of individuals with AD. They justify the Markov property by citing studies which have found no association between duration of AD and increased mortality (Barclay et al., 1985; Bracco et al., 1994; Burns et al., 1991; Diesfeldt et al., 1986; Heyman et al., 1996; Sayetta, 1986; Walsh et al., 1990).

In the absence of data to provide a more suitable model, and for mathematical convenience, we assume all state transitions depend only on the current state, *i.e.* are Markov.

For each state ik, we call i its cognitive ability type (the progress of dementia) and k its functional ability type (the number of IADL or ADLs). Changes in cognitive ability type are assumed to be independent of functional ability type. Similarly, changes in functional ability type are assumed to be independent of cognitive ability type. To express this mathematically, we have $\mu_{x+t}^{ik,il} = \mu_{x+t}^{jk,jl}$ and $\mu_{x+t}^{ik,jk} = \mu_{x+t}^{il,jl}$.

McGuire et al. (2006) suggest the latter of these assumptions is not valid. In an American population, they found those with the lowest level of cognition (as diagnosed by telephone interview) had greater odds of ADL disability. Iwasa et al. (2008) conducted an analysis of how specific areas of cognitive function related to functional decline among a Japanese community. Their results showed a link between ADLs and either a decline in information processing speed or orientation to time and place. However due to different methods in diagnosis of cognitive function and functional ability among studies, reliable data for the purposes of fitting into our model is scarce. We therefore continue with this assumption in place and now parameterise the transitions between states.

Inclusion of recoveries adds complexity of the model. However as mentioned above, Pritchard (2006) found that excluding recoveries results in overstating the cost of benefits. For this reason, in our model it is possible to recover from functional disabilities.

3.2.1 Notation

We introduce the notation which will be used in this section:

1. As introduced above (and included here for ease of reference), $\mu_{x+t}^{ik,jl}$ is the transition intensity from state ik to state jl at age $x+t$.

2. In the case dementia, where genotype specific intensities are necessary, denote the transition intensity from state, ik, to state, jl, at age, $x+t$, for a life with genotype, g, by $\mu_{x+t,g}^{ik,jl}$.

3. The relative risk of AD cause by APOE genotype, g, at age, x, is denoted by, $\varrho_{x,g}$.

4. Let $\hat{\mu}_{x+t}^{i}$ be the ungraduated force of mortality at age $x+t$ from state $0i$ estimated by Akodu (2007).

5. The graduated relative risk of mortality caused by functional disability type i, relative to lives with no ADLs, is denoted by ρ_i; while ungraduated for lives aged x, is denoted by $\hat{\rho}_{x,i}$.

6. In common with standard actuarial notation, q_{x+t} and p_{x+t} are the one year probabilities of death and survival respectively, for a life aged $x+t$. After applying mortality improvements, these become $q_x t$ and $p'_{x,t}$ respectively.

7. The factor by which to reduce the 1-year probability of death, t years after the date of the underlying mortality table for a life who was aged x at $t=0$ is denoted by $RF_{x+t,t,y}$. The The associated reduction factor applicable to $\mu_{x+t}^{00,4}$, is denoted by $v_{x,t}$.

8. The set of all states in our final Markov model of health is denoted by $\mathcal{S}$.

3.2.2 Functional Ability

We first consider the transition between functional ability types. Transitions between functional ability types are reversible, *e.g.* it is possible that a person could go from Healthy to > 1 ADL and subsequently 'recover' to having 1 ADL. Transition intensities are therefore required in both directions connecting each functional ability type, for a given cognitive function type.

Functional ability transition intensities are taken to be those derived by Akodu (2007), who fitted functions to using the penalised least squares method of Pritchard (2006). This is the only study that calculates age dependent incidence of functional ability for a U.K. population. It is based on a large body of data (13,004 lives in the first phase of a two phase, longitudinal study) and is fitted with a Markov model, consistent with our assumptions.

A limited number of other studies report the incidence of functional disability but these are usually concerned with predictive factors for its onset and do not include recovery, *e.g.* Alexandre et al. (2012) found age dependent incidence by particular ADLs for a Brazilian population; and Taş et al. (2007) performed a prospective longitudinal study from the Rotterdam Study (Hofman et al., 1991). Pritchard (2002) included transitions between various levels of functional disability (defined by number of ADLs) using the US National Long-Term Care Surveys.

Akodu (2007)'s intensities were fitted using the penalised maximum likelihood method of Pritchard (2006) and graduated to the following functions depending on which best fit the shape:

- Makeham: $\mu_{x+t}^{ik,il} = \alpha_{kl} + \beta \exp\left(\gamma_{kl}\left(x - 68.5 + t\right)\right)$;

- Weibull: $\mu_{x+t}^{ik,il} = \frac{\gamma_{ij}}{\beta_{kl}}\left(\frac{x+t}{\beta_{ij}}\right)^{\gamma_{kl}-1}$; and

- Linear: $\mu_{x+t}^{ik,il} = \alpha_{kl} + \beta_{ij}\left(x+t\right)$.

Tables 3.1 and 3.2 list where each applies and their corresponding parameters.

3.2.3 Mortality Without Dementia

We next parameterise the mortality for lives who have not developed dementia (*i.e* from state $0j$). We observe in Akodu (2007) that, apart from lives with 1 ADL, the mortality rates experienced by lives in functionally disabled states are significantly higher than the point estimates for lives with no ADLs. Therefore we will need to fit mortality dependent on functional ability. For simplicity we will use a relative risk approach, where we have some base mortality rate applicable to lives who have

Table 3.1: Parameters and functions for transition intensities between functional ability type states for males. Source: Akodu (2007).

From Status (k)	To Status (l)	Parameter			Fitted Function
		α_{kl}	β_{kl}	γ_{kl}	
None	IADL	-	7.3×10^1	3.89×10^0	Weibull
	1-ADL	1.72×10^{-2}	5.24×10^{-3}	1.63×10^{-1}	Makeham
	≥ 2 ADLs	-1.68×10^{-2}	2.99×10^{-4}	-	Linear
IADL	None	9.51×10^{-1}	-9.66×10^{-3}	-	Linear
	1-ADL	9.6×10^{-2}	1.01×10^{-3}	-	Linear
	≥ 2ADLs	-1.13×10^{-2}	1.26×10^{-2}	1.21×10^{-1}	Makeham
1-ADL	None	6.54×10^{-1}	-6.76×10^{-3}	-	Linear
	IADL	4.38×10^{-2}	1.60×10^{-5}	-	Linear
	≥ 2ADLs	2.81×10^{-1}	-1.44×10^{-3}	-	Linear
≥ 2ADLs	None	-5.58×10^{-2}	8.65×10^{-4}	-	Linear
	IADL	5.74×10^{-2}	-3.89×10^{-4}	-	Linear
	1-ADL	-7.12×10^{-2}	2.30×10^{-3}	-	Linear

Table 3.2: Parameters and functions for transition intensities between functional ability type states for females. Source: Akodu (2007).

From Status (k)	To Status (l)	Parameter			Fitted Function
		α_{kl}	β_{kl}	γ_{kl}	
None	IADL	-4.95×10^{-1}	8.35×10^{-3}	-	Linear
	1-ADL	-	7.64×10^1	5.22×10^0	Weibull
	≥ 2ADLs	1.54×10^{-2}	-1.70×10^{-4}	-	Linear
IADL	None	6.17×10^{-1}	-5.48×10^{-3}	-	Linear
	1-ADL	-6.28×10^{-1}	1.09×10^{-2}	-	Linear
	≥ 2ADLs	8.53×10^{-2}	1.78×10^{-3}	-	Linear
1 ADL	None	4.89×10^{-1}	-5.18×10^{-3}	-	Linear
	IADL	-1.22×10^{-1}	2.18×10^{-3}	-	Linear
	≥ 2ADLs	-	6.36×10^1	4.52×10^0	Weibull
≥ 2ADLs	None	2.88×10^{-2}	-3.19×10^{-4}	-	Linear
	IADL	2.83×10^{-2}	-2.96×10^{-4}	-	Linear
	1 ADL	5.25×10^{-1}	-4.87×10^{-3}	-	Linear

no functional disability and apply some factor to calculate mortality for lives with a functional disability.

For a level of consistency with Macdonald and Pritchard (2001), we choose as our base mortality rate for lives with no cognitive disability, the CMI (2009)'s AMC00 and AFC00 assured lives mortality rates for male and female lives respectively (the previous study used CMI (1990)'s AM80 and AF80, previous versions of assured mortality). They were calculated from the experience of subscribing U.K. insurance companies over the period 1999–2002, and they apply to a life attaining age x on the 1st of July, 2000.

In this most recent series of mortality tables, there are mortality rates dependent on smoker status. Although it is common practice for insurance companies to use separate tables in their modelling, we use the combined tables. There is evidence that smoking is a significant factor for increasing the risk of developing AD (Cataldo et al., 2010). However, if we ignore fraudulent non-disclosure, information of smoker status is not hidden from the insurer — insurers routinely ask on the application form — so there is no information asymmetry and consequently no adverse selection. We therefore feel the choice to use the combined tables is justified in keeping the complexity of the model to a minimum.

Both sets of mortality data were graduated by the CMI using Gompertz-Makeham formulae for ages $x \leq 100$ and a blending formula to blend the rate from age $x = 100$, μ_{100}, to $\mu_{120} = 1$ with specified curve parameter. For males they use

$$\mu_x^{00,4} = \begin{cases} \exp\left[-4.01872 + 5.8902\frac{x-70}{50} - 1.1515\left(\frac{x-70}{50}\right)^2\right] \\ +0.00044726 & \text{if } 17 < x \leq 100, \\ 0.407367\left(\frac{(120-x)^{1.25}}{20^{1.25}}\right) + \left(1 - \frac{(120-x)^{1.25}}{20^{1.25}}\right) & \text{if } 100 < x \leq 120, \end{cases}$$

and for females they use

$$\mu_x^{00,4} = \begin{cases} 0.00014423 + \exp\left(-4.389068 + 5.584346\frac{x-70}{50}\right) & \text{if } 17 < x \leq 100, \\ 0.354144\left(\frac{(120-x)^{1.25}}{20^{1.25}}\right) + \left(1 - \frac{(120-x)^{1.25}}{20^{1.25}}\right) & \text{if } 100 < x \leq 120. \end{cases}$$

Now we calculate the relative risks of death while functionally disabled using data from Akodu (2007). Akodu (2007) performed analysis on the CFAS study which sampled lives from sites around the U.K., over the period 1991–1994. We assume these relative risks are applicable to our chosen mortality rates.

The ungraduated relative risk of mortality, relative to lives with no ADLs, $\hat{\rho}_{x,i}$, is calculated as

$$\hat{\rho}_{x,i} = \frac{\hat{\mu}_x^i}{\hat{\mu}_x^0}. \tag{3.1}$$
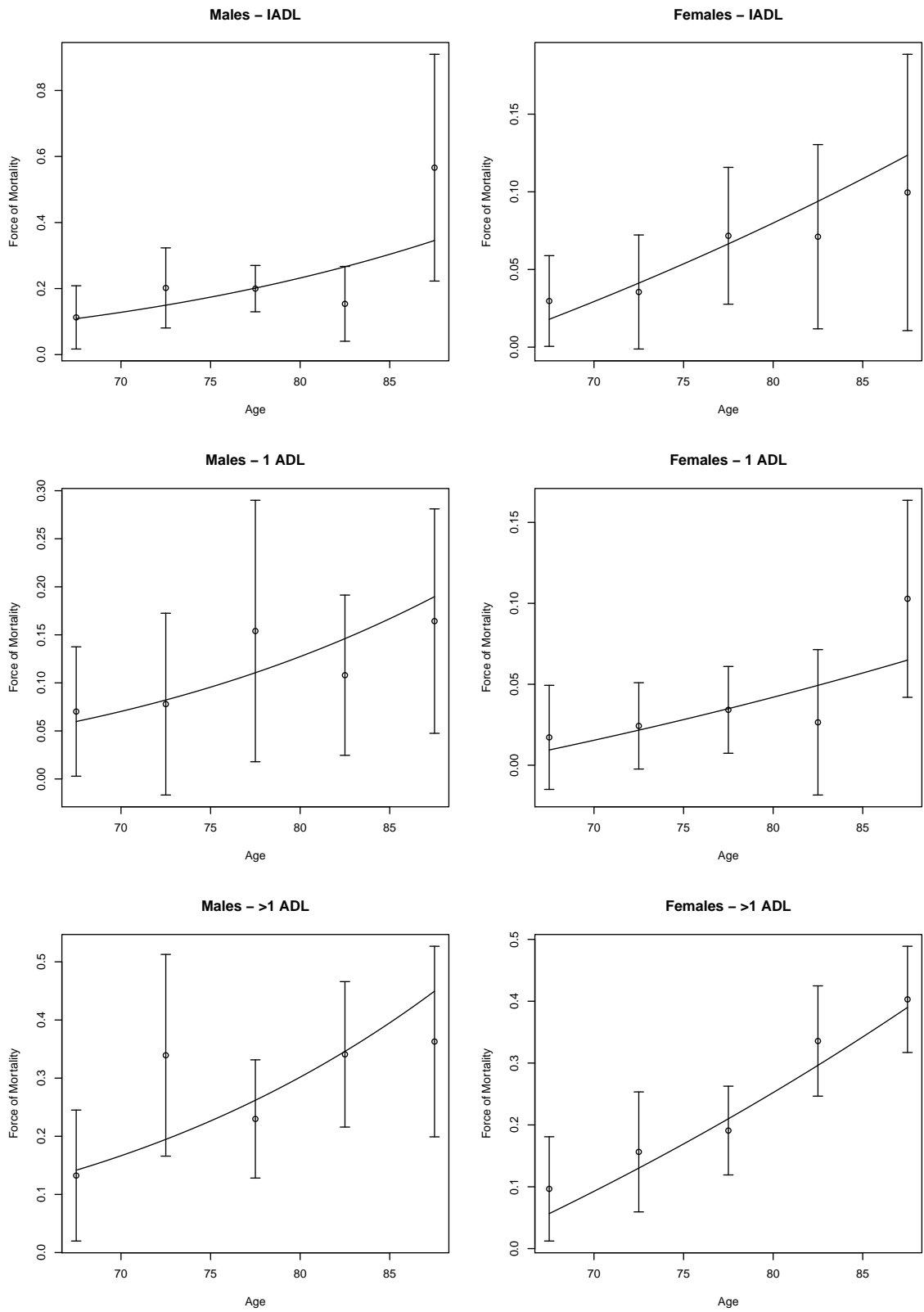
Figure 3.2: Forces of mortality in a relative risk model. The force of mortality for lives with no disability, point estimates and confidence intervals are from Akodu (2007).

Table 3.3: Relative risk of mortality according to functional ability type, relative to a life with no ADLs, for males and females.

Functional Ability	Male	Female
IADL	5.179	2.770
1-ADL	2.844	1.456
$\geq$ 2ADLs	6.737	8.746

For simplicity, we fit a constant relative risk of mortality for functional disability type i, as the weighted average over all ages, where the weight for age x is $1/Var(\hat{\mu}_x^i)$. These are shown in Table 3.3. The mortality rate for state ij is given by

$$\mu_{x+t}^{ij,4} = \rho_i \mu_{x+t}^{0j,4}. \tag{3.2}$$

To check consistency with the CFAS data, we apply these relative risks, relative to Akodu (2007)'s fitted force of mortality for lives with no ADLs, in Figure 3.2. For each level of functional disability, the force of mortality is within 95% confidence intervals at every age, hence there would be no evidence against our constant relative risks.

3.2.4 Mortality Improvements

Recent years have shown substantial improvements in mortality (Office for National Statistics, 2012b). Willets et al. (2004) suggests that over the past century, the biggest driver of these improvements was the development of vaccines and antibiotics to treat infectious disease. They also outline the areas causing mortality change at the start of this century:

- The 'cohort effect' — the phenomenon whereby mortality improvement differs by year of birth. In particular those born in the U.K. during the period 1925–1945 have experienced the most rapid improvements in mortality compared to other generations.

- The 'ageing of mortality improvement' — over time, the ages which display the highest rates of mortality improvement are increasing.

- Increased uncertainty at younger ages.

- Changes in prevalence of cigarette smoking.

- Widening social class differentials — mortality has improved more quickly for those in higher socio-economic classes.

Without projecting changes in mortality in a pricing basis, insurance contracts that extend over a long period could experience have significantly lighter mortality than what was assumed. For products which provide benefits contingent on survival, this could mean longer lasting periods of benefit payments than had been priced for, potentially causing a loss. LTC is one such product — benefits are paid while the insured is alive in a disabled state. We therefore include allowance for mortality improvement in our modelling, but we assume no changes to the relative risks associated with functional disability calculated above.

To project the mortality improvements, we use CMI (2011)'s mortality projections model, which combines cohort effects and age effects of mortality improvement. This model uses a large number of age and birth year dependent assumptions, which are listed in Appendix C.

Each component is projected independently from the year 2008. The method uses a cubic formula to blend the most recently observed improvement rate (which we will refer to as the initial rate) to some long-term average, over an assumed period of years. The cohort component is assumed to tend to zero in the long-term. For the age-effect, we choose a long-term average of 1% at all ages, for males and females; this is consistent with the Government Actuary's Department's principle projection[1]. The rate of convergence is controlled by a parameter, which we denote by β, for the level of convergence achieved at the midpoint of the blending period. Let $w\left(\beta, t, T\right)$, be the weighting of the most recently observed improvement rate for the projected mortality improvement at time t, during a convergence period lasting T years. This is calculated as,

$$w\left(\beta, t, T\right) = a\left(\frac{t}{T}\right)^3 + b\left(\frac{t}{T}\right)^2 + \frac{t}{T} + 1, \tag{3.3}$$

where $a = -2 + 8\beta$, $b = 5.16\beta$ and $c = -4 + 8\beta$.

Let $\zeta_{x,t}^{\text{Age}}$ be the age-effects component of the mortality improvement rate for a life aged x, at time, t. $\zeta_{x,0}^{\text{Age}}$ and $\zeta_{x,T_x^{\text{Age}}}^{\text{Age}}$ therefore represent the age effect initial and long-term rates respectively. $\zeta_{x,t}^{\text{Age}}$ is calculated as,

$$\zeta_{x,t}^{\text{Age}} = w\left(\beta, t, T_x^{\text{Age}}\right)\zeta_{x,0}^{\text{Age}} + \left(1 - w\left(\beta, t, T_x^{\text{Age}}\right)\right)\zeta_{x,T_x^{\text{Age}}}^{\text{Age}}, \tag{3.4}$$

where T_x^{Age} is the period of convergence for the age-effects component at age x. The cohort effect component is calculated similarly.

The projected mortality improvement rate for a life aged x in the year $2008 + t$, is denoted by $\zeta_{x,t}$ and calculated by combining the age-effect and cohort components:

$$\zeta_{x,t} = \zeta_{x,t}^{\text{Age}} + \zeta_{2008+t-x,t}^{\text{Cohort}}. \tag{3.5}$$

---

[1] `http://www.gad.gov.uk/Demography\%20Data/Life\%20Tables/Varmortass.html`, retrieved 31/07/2013.

For non-positive values of t, *i.e.* for year 2008 and earlier, we use the experienced mortality improvement rates, which are listed in Appendix C.

The method used by Continuous Mortality Investigation (2011) to apply these improvement rates to a mortality table, is to first calculate a set of base reduction factors, using all of the available observed data, *i.e.* from 1st January, 1990. From this base reduction factor, they rebase it to the date of the mortality table being used. The base reduction factors, $RF^*_{x,t}$ are calculated at integer age x, and time since the date of our mortality table, t, as,

$$RF^*_{x,t} = \begin{cases} RF^*_{x,t-1}\left(1 - \zeta_{x,1990-2008+t-1}\right) & \text{if } t \in \mathbb{Z}^+ \\ 1 \text{ if } t = 0. \end{cases} \tag{3.6}$$

From these base reduction factors, the reduction factors for a mortality table applicable y years after 1st January, 1990, $RF_{x,t,y}$ are calculated as,

$$RF_{x,t,y} = \frac{RF^*_{x,t+\lfloor y \rfloor}{}^{1-(y-\lfloor y \rfloor)} RF^*_{x,t+\lfloor y \rfloor+1}{}^{y-\lfloor y \rfloor}}{RF^*_{x,\lfloor y \rfloor}{}^{1-(y-\lfloor y \rfloor)} RF^*_{x,\lfloor y \rfloor+1}{}^{y-\lfloor y \rfloor}}. \tag{3.7}$$

where x and t are integers. For non-integer x and t, we calculate $RF_{x,t,y}$ using geometric interpolation: first on age, and secondly on time, as

$$\begin{aligned} RF_{x,t,y} &= \left(RF_{\lfloor x \rfloor,\lfloor t \rfloor,y}{}^{1-(x-\lfloor x \rfloor)} RF_{\lfloor x \rfloor+1,\lfloor t \rfloor,y}{}^{x-\lfloor x \rfloor}\right)^{1-(t-\lfloor t \rfloor)} \\ &\quad \times \left(RF_{1+\lfloor x \rfloor,1+\lfloor t \rfloor,y}{}^{1-(x-\lfloor x \rfloor)} RF_{2+\lfloor x \rfloor,1+\lfloor t \rfloor,y}{}^{x-\lfloor x \rfloor}\right)^{(t-\lfloor t \rfloor)}. \end{aligned} \tag{3.8}$$

This produces factors to reduce q_x, the 1-year probability of death, at integer ages and times: $q'_{x,t} = RF_{x+t,t,y}\, q_{x+t}$ and $p'_{x,t} = 1 - q'_{x,t}$. However, we use the force of mortality, $\mu^{00,4}_{x+t}$ so we need to calculate a consistent factor to improve the force of mortality.

Now, $RF_{x,0,y} = 1$, *i.e.* $q_x = q'_{x,0}$, therefore $\int_0^1 \mu_{x+t}^{00,4} \, dt = \int_0^1 \upsilon_{x,t} \, \mu_{x+t}^{00,4} \, dt$ so $\upsilon_{x,t} = 1$ for $0 \leq t \leq 1$. For $t \geq 1$ we have

$$\frac{_h p'_{x,t}}{_h p'_{x,t-1}} = \frac{p'_{x,t+h-1}}{p'_{x,t-1}}$$

$$_h p'_{x,t} = \frac{p'_{x,t+h-1} \; _h p'_{x,t-1}}{p'_{x,t-1}}$$

$$= \frac{_{1+h} p'_{x,t-1}}{p'_{x,t-1}}$$

$$\exp\left(-\int_0^h \upsilon_{x,t+s} \, \mu_{x+t+s}^{00,4} \, ds\right) = \frac{p'_{x+t+h-1} \, \exp\left(-\int_0^h \upsilon_{x,t-1+s} \, \mu_{x+t-1+s}^{00,4} \, ds\right)}{p'_{x,t-1}}$$

$$-\int_0^h \upsilon_{x,t+s} \, \mu_{x+t+s}^{00,4} \, ds = \log\left[\frac{p'_{x,t+h-1}}{p'_{x,t-1}} \exp\left(-\int_0^h \upsilon_{x,t-1+s} \, \mu_{x+t-1+s}^{00,4} \, ds\right)\right]$$

$$\upsilon_{x,t+h} \, \mu_{x+t+h}^{00,4} = -\frac{\frac{d}{dh} p'_{x,t+h-1}}{p'_{x,t+h-1}} + \upsilon_{x,t-1+h} \, \mu_{x+t-1+h}^{00,4}$$

$$\upsilon_{x,t+h} = \frac{-\frac{\frac{d}{dh} p'_{x,t+h-1}}{p'_{x,t+h-1}} + \upsilon_{x,t-1+h} \, \mu_{x+t-1+h}^{00,4}}{\mu_{x+t+h}^{00,4}}. \tag{3.9}$$

For $t \geq 1$, to solve Equation (3.9), we approximate the derivative $\frac{d}{dh} p'_{x,t+h-1}$ numerically with a central difference method,

$$\frac{d}{dh} p'_{x,t+h-1} \approx \frac{p'_{x,t+2h-1} - p'_{x,t-1}}{2h}. \tag{3.10}$$

The mortality intensities from the healthy state are for lives attaining age x on 1st of July, 2000, while our modelling date is some 12.5 years later, on the 1st of January, 2013. The results of this mortality improvement applied to lives in Healthy are shown in Figure 3.3. The force of mortality used by Macdonald and Pritchard (2001) is higher than what we use, particularly for males; while at the oldest ages theirs is lower, particularly for females. This is driven by the pattern of mortality improvements: as we have discussed, there are birth year and age effects on mortality improvement, whereas Macdonald and Pritchard (2001) assumed an average improvement factor applicable to all lives.

A noticeable feature of the improved mortality intensities is that they are not continuous functions. The reason is the recursive nature of calculating $\upsilon_{x,t}$. We could write Equation (3.9) as a system of 2^h non-homogeneous recurrence relations, denoted
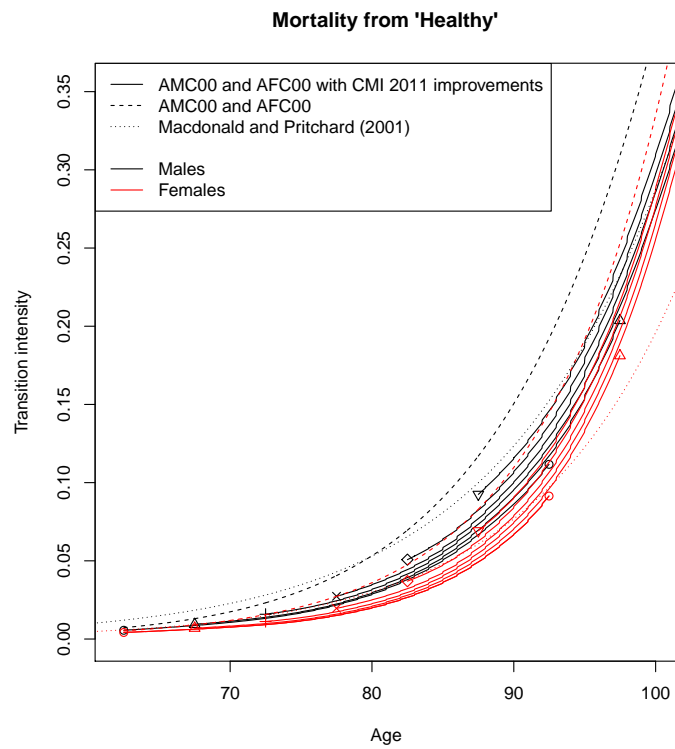
Figure 3.3: Force of mortality from the 'Healthy' state with and without mortality improvements. Improved mortality shown from 2013 with points representing the start and end points of a 30 year period for lives aged $62.5, 67.5, \ldots, 87.5$ in 2013.

by $f_{x,s,n}$, with $0 < s \leq 1$ and $s + n = t > 1$, in the form:

$$v_{x,t} = f_{x,s,n} = g(x, s + n) + f_{x,s,n-1}\,\mu(x + s + n), \qquad (3.11)$$

where $g(x, s+n) = -\frac{d}{ds}p'_{x,s+n-1}/(p'_{x,s+n-1}\mu^{00,4}_{x+s+n})$, and $\mu(x+s+n) = \mu^{00,4}_{x+s+n-1}/\mu^{00,4}_{x+s+n}$ and $f_{x,s,0} = 1$. To be continuous at $t = 1$ we need $\lim_{t \to 1^+} v_{x,t} = 1$, but at age $x = 62.5$ for instance, we have

$$\lim_{t \to 1^+} v_{62.5,t} = \frac{\frac{-\frac{d}{ds}p'_{62.5,s-1}\big|_{s=1}}{p'_{62.5,0}} + \mu^{00,4}_{62.5}}{\mu^{00,4}_{63.5}} \approx 0.8607. \qquad (3.12)$$

Hence, $v_{62.5,t}$ is discontinuous at $t = 1$, and due to the recurrence relationship, is then discontinuous at each integer value of t. Since the functions, $g(x, s + n)$ and $\mu(x + s + n)$ are continuous at non-integer values of $t = s + n$, we have the piece-wise continuity observed.

3.2.5 Genetics of Alzheimer's Disease

In this section we describe the impact of the APOE gene in AD onset and parameterise a genetic component of the disease into our model.

The major genetic risk factor for AD is the APOE gene (Coon et al., 2007), which codes for a protein that transports lipids around the circulatory system. There are at least three allele variants, the most common of which are named $\varepsilon 2$, $\varepsilon 3$ and $\varepsilon 4$. This leads to possible 6 genotypes: $\varepsilon 2\varepsilon 2$, $\varepsilon 2\varepsilon 3$, $\varepsilon 2\varepsilon 4$, $\varepsilon 3\varepsilon 3$, $\varepsilon 3\varepsilon 4$ and $\varepsilon 4\varepsilon 4$. Relative to $\varepsilon 3$, $\varepsilon 4$ makes a life more susceptible to developing AD, while $\varepsilon 2$ reduces susceptibility. However, the processes that cause these differences in susceptibility are yet to be established (Huang and Mucke, 2012).

It is possible to test which variants are present in a life's genome, indeed this is among the genes tested by companies offering genetic tests to individuals such as 23andMe[2]. However it is unlikely to be used predictively by doctors in relation to AD since there is currently no preventative treatment available and the tests results might create undue stress to the life. Low et al. (2010) found there was insufficient evidence to recommend testing to aid in prescribing risk reduction methods for dementia, while Farrer et al. (1995) recommends against its use for the reason that APOE alone is insufficient in accurately predicting age of onset.

APOE was considered to be a susceptibility gene for coronary heart disease (CHD) (Song et al., 2004). However, there is debate as to the validity of the link. In a large study which controlled for the a variety of known cardiovascular risk factors, Ward et al. (2009) found no association between APOE and CHD. A recent study

---

[2]`https://www.23andme.com/health/all/`

by Kofler et al. (2012) analysed the interaction of APOE with body mass index, age and sex and concluded that the risk of developing CHD due to APOE genotype was 'unlikely to be homogeneous'. Nonetheless, while the link and risk factors of particular genotypes for CHD are still being studied, it remains plausible that testing for APOE genotype might one day be useful in devising treatment for diseases other than AD. This broadens the possibility that individuals could learn of their genetic susceptibility to AD even if that was not the intention. We therefore incorporate APOE genotypes into our model of health which will allow us to use this as a source of adverse selection.

In actuarial studies, age dependent relative risks are ideal for fitting genotype specific transition rates, however this is rarely available from epidemiological studies. Macdonald and Pritchard (2000) used results from Farrer et al. (1997) to fit age dependent relative risks. As they noted however, the data for males was sparse, particularly for older ages. Consequently, the pattern seen at older ages — where $\varepsilon 2\varepsilon 4$ and $\varepsilon 3\varepsilon 4$ have lower risk relative to $\varepsilon 3\varepsilon 3$ in males — was considered possibly due to the lack of data since females showed higher risk in all $\varepsilon 4$ variants. It is possible that the biological effect of the $\varepsilon 4$ allele on the oldest males is to reduce relative risk, however Macdonald and Pritchard (2001) made an adjustment to their previous genetic model, such that $\varepsilon 2\varepsilon 4$, $\varepsilon 3\varepsilon 4$ and $\varepsilon 4\varepsilon 4$ are restricted to be at least the relative risk of $\varepsilon 3\varepsilon 3$ in males.

Denote the set of APOE genotypes by $\mathcal{G}$. Due to the rarity of $\varepsilon 2\varepsilon 2$ and consistent with Macdonald and Pritchard (2000) and Macdonald and Pritchard (2001), we group these together with $\varepsilon 2\varepsilon 3$ and consider them indistinguishable. Therefore we have $\mathcal{G} = \{\varepsilon 2\varepsilon 2, \varepsilon 2\varepsilon 4, \varepsilon 3\varepsilon 3, \varepsilon 3\varepsilon 4, \varepsilon 4\varepsilon 4\}$.

We use the fitted relative risks of Macdonald and Pritchard (2001) because these are still the most useful available. For a life aged $x + t$ with APOE genotype $g \in \mathcal{G}$, these are of the form,

$$\varrho_{x+t,g} = \begin{cases} r_m \left[\left(\max \left(f_{x+t}^g, 1 \right) - 1 \right) m + 1 \right] & \text{if male and } g \in \{\varepsilon 2\varepsilon 4, \varepsilon 3\varepsilon 4\} \\ r_m \left[\left(f_{x+t}^g - 1 \right) m + 1 \right] & \text{otherwise,} \end{cases} \tag{3.13}$$

where m, is a parameter to reduce the relative risks, allowing for ascertainment bias and

$$f_{x+t}^g = E \exp \left[-F \left(x + t - k_1 \right)^2 - G \left(x + t - k_2 \right) \right] + H. \tag{3.14}$$

The fitted parameters are in Table 3.4.

Table 3.4: Parameters for use with the $GM(1,3)$ function in Equation 3.14 for the transition intensities of AD for males and females by genotype. Source: Macdonald and Pritchard (2001)

Gender	Genotype	E	F	G	H	k_1	k_2	r_1
Female	$\varepsilon 4\varepsilon 4$	10.400	0.00504	0	1	60	0	0.88
	$\varepsilon 3\varepsilon 4$	3.680	0.00319	0	1	62	0	
	$\varepsilon 2\varepsilon 4$	4.210	0.01020	0	1	68	0	
	$\varepsilon 2\varepsilon 2$	0.675	0	0.00692	0	0	60	
Male	$\varepsilon 4\varepsilon 4$	8.940	0.00656	0	1	60	0	1.12
	$\varepsilon 3\varepsilon 4$	1.920	0.00103	0	0	51	0	
	$\varepsilon 2\varepsilon 4$	1.420	0.00506	0	0	67	0	
	$\varepsilon 2\varepsilon 2$	0.434	0	0.01600	0	0	60	

3.2.6 Onset of Dementias

In this section we will fit transition intensities for non-AD and AD respectively, in order to apply gene specific relative risks only to AD onset. Our model does not have separate states for the different types of dementia, so we calculate dementia onset as the total of these two components.

Launer et al. (1999) pooled 4 population-based prospective studies of onset rates and risk factors for dementia as a whole and AD in particular by Letenneur et al. (1994); Copeland et al. (1999); Andersen et al. (1999); Ott et al. (1995). These were all part of the European Studies of Dementia (EURODEM) Launer (1992). Being part of the EURODEM network, each of these studies were performed with harmonised protocols: Participants were either randomly selected or were all of the elderly in a geographic area. After initial screening where anyone found to have dementia was excluded, follow-up screening was after approximately 2 years. Each study used American Psychiatric Association (1987) criteria to diagnose mild to severe dementia. In particular, all patients were given MMSE, the organic section of the Geriatric Mental State Examination and the Cambridge Examination of Mental Disorders Cognitive Test. Where the scores of these tests were below a cutoff point, more detailed diagnostic testing took place. This standardisation of methodology allows the pooling of the results.

We choose Launer et al. (1999) because they reported exposures and number of diagnoses for AD and dementia in 5-year age groups which are shown in Table 3.5, allowing maximum likelihood estimation of model parameters. Moreover, the largest contribution to EURODEM, by size of cohort, came from the U.K. study, Copeland et al. (1999).

For their onset of AD, Macdonald and Pritchard (2001) fitted a Gompertz curve using a log-linear least squares method to the point estimates of AD incidence by

Table 3.5: Results from a pooled analysis of studies in the EURODEM network. Source: Launer et al. (1999).

| Age | Non-AD | | AD | |
	Exposure e_x^{NAD}	No. of Diagnoses d_x^{NAD}	Exposure e_x^{AD}	No. of Diagnoses d_x^{AD}
65–69	6352	6	6340	7
70–74	7778	17	7755	21
75–79	6529	43	6462	63
80–84	4538	38	4489	97
85–89	2390	39	2341	89
≥ 90	1181	33	1144	75

Jorm and Jolley (1998). Jorm and Jolley (1998) collected the results of 23 studies, 13 of them European. Of these European studies, only 5 included the mild classification of AD used by Macdonald and Pritchard (2001). Exposure values were not included so maximum likelihood estimates were not available.

The genetic relative risks for AD in Section 3.2.5 are sex specific and applied to the aggregate onset rate so we fit the aggregate onset rate here. However, we note that although results separated by gender were not published, sex was found to be a significant risk factor for AD by Launer et al. (1999), with female rates 54% higher than male rates, but no significant difference was observed for non-AD dementia.

Denote the population average onset rate of AD at age x by μ_x^{AD} and the onset rate of non-AD dementia by μ_x^{NAD}. Due to the small size of the data set, we seek the simplest relationship between non-AD dementia and AD dementia transition intensities such that $\mu_x^{NAD}/\mu_x^{AD} = \kappa$ for some constant κ. We first check heuristically whether this is a viable option by considering the confidence intervals of these ratios. For each age x, we use Graham et al. (2003)'s method of constructing lower and upper confidence limits of the ratio of Poisson rates, denoted $\phi_L(x)$ and $\phi_U(x)$ respectively:

$$\phi_L(x) = \frac{e_x^{AD}}{e_x^{NAD}} \left[\frac{2d_x^{AD}d_x^{NAD} + z_\alpha^2 d_x - \sqrt{z_\alpha^2 d_x \left(4d_x^{AD}d_x^{NAD} + z_\alpha^2 d_x\right)}}{2d_x^{AD2}} \right] \quad (3.15)$$

$$\phi_U(x) = \frac{e_x^{AD}}{e_x^{NAD}} \left[\frac{2d_x^{AD}d_x^{NAD} + z_\alpha^2 d_x + \sqrt{z_\alpha^2 d_x \left(4d_x^{AD}d_x^{NAD} + z_\alpha^2 d_x\right)}}{2d_x^{AD2}} \right], \quad (3.16)$$

where z_α is the inverse of the cumulative distribution function of a standard normal distribution evaluated at α. The results of this are shown in Table 3.6 for $\alpha = 0.025$ *i.e.* a 95% confidence interval.

For values of κ such that $\sup\{\phi_L(x); 65 < x\} \leq \kappa \leq \inf\{\phi_U(x); 65 < x\}$, the ratio $\mu_x^{NAD}/\mu_x^{AD} = \kappa$ will be within the 95% confidence intervals at every age x. Hence

Table 3.6: Confidence limits for the ratio of μ^{NAD}/μ^{AD}.

Age	ϕ_L	ϕ_U
65–69	0.302	2.427
70–74	0.430	1.514
75–79	0.460	0.993
80–84	0.267	0.563
85–89	0.295	0.624
≥ 90	0.284	0.640

there will be insufficient evidence to reject a constant ratio of the transition intensities of AD and non-AD dementias for such values of κ.

Using the method of maximum likelihood, we fit the the Gompertz-Makeham (GM) family of models to μ_x^{AD} and μ_x^{NAD}. If function $f(x)$ is GM(r,s), then

$$f(x) = \sum_{i=0}^{r} a_i\, x^{i-1} + \exp\left(\sum_{i=0}^{s} b_i\, x^{i-1} \right), \tag{3.17}$$

for some values a_i, b_j, $i \leq r, j \leq s$ and $a_0 = b_0 = 0$.

With $\mu_x^{NAD} = \kappa\, \mu_x^{AD}$, the likelihood function to be maximised is

$$L \propto \frac{\left(e_x^{NAD}\, \kappa\, \mu_x^{AD}\right)^{d_x^{NAD}}}{d_x^{NAD}!} \frac{\left(e_x^{AD}\, \mu_x^{AD}\right)^{d_x^{AD}}}{d_x^{AD}!} \exp\left(-e_x^{NAD}\, \kappa\, \mu_x^{AD} - e_x^{AD}\, \mu_x^{AD}\right). \tag{3.18}$$

The results of this maximisation are shown in Table 3.7 along with the log of the likelihood attained and the Akaike information criterion (AIC).

Many of the fitted models are nested — models are said to be nested when one can be considered to be a more generalised form of another *e.g.* GM(0,2) and GM(0,3) are nested. We fit each model independently, so our nesting requires additional parameters over those implied by Equation (3.17) to allow for difference in parameter values. Taking our example, we write a GM(0,2) as a GM(0,3) in the form $\mu_x^{NAD} = (\kappa + \kappa') \exp[b_1 + b_1' + (b_2 + b_2')x + b_3'x^2]$, where κ, b_1 and b_2 are the fitted values when maximising the likelihood for GM(0,2).

We perform likelihood ratio tests on nested models to test whether the addition of extra parameters in the alternative model provides a significantly better fit to the data. In the likelihood ratio test, the null hypothesis is that the additional parameters required to express as the generalised form are equal to zero. Under the null hypothesis, the test statistic $D = 2\,(l_a - l_0)$, where l_a and l_0 are the log likelihoods under the alternative and null models respectively, is approximately distributed with a χ^2-distribution, with degrees of freedom equal to the number of parameters set equal to zero (Mendenhall, 1986). In our example, we have GM(0,2) as the null model and

GM(0,3) as the alternative and the null hypothesis is that $\kappa' = b'_1 = b'_2 = b'_3 = 0$, hence four degrees of freedom. The results of this testing are shown in Table 3.8.

We find that GM(0,3) and GM(1,2) are candidates for the best model and we choose the one with the smallest AIC which is the GM(0,3).

$$\mu_x^{NAD} = 0.5\mu_x^{AD} = 0.5\exp\left(-40.9847 + 0.762795x - 0.00379x^2\right). \tag{3.19}$$

The overall dementia intensity is therefore,

$$\mu_{x,g}^{0i,1i} = (0.5 + \varrho_{x,g})\exp\left(-40.9847 + 0.762795x - 0.00379x^2\right). \tag{3.20}$$

The fitted transition intensities for AD and overall dementia (with AD relative risk, $\varrho_{x,g} = 1$) are shown in Figure 3.4 with the confidence intervals around the point estimates of Launer et al. (1999). The fitted models of Macdonald and Pritchard (2001) and Akodu (2007) are also shown.

Comparing the AD model of Macdonald and Pritchard (2001) with the data of Launer et al. (1999), we see that their model produces transition intensities slightly (but significantly) larger. There is great variability in studies of AD, which Corrada et al. (1995) found to be partially due to differences in methodology. It is possible that differences in methodology exist in the various studies used in the meta-analysis of Jorm and Jolley (1998) and those pooled by Launer et al. (1999), contributing to the differences observed here.

Table 3.7: Fitted parameters of Gompertz-Makeham type models to onset of dementia.

Model	a_1	a_2	a_3	b_1	b_2	b_3	b_4	b_5	κ	Log Likelihood	AIC
(0,2)	-	-	-	-15.5270	0.1394	-	-	-	0.4916	-2693.01	5392.03
(0,3)	-	-	-	-40.9847	0.7628	-0.0038	-	-	0.5000	-2687.20	5382.41
(0,4)	-	-	-	-40.9847	0.7628	-0.0038	0.0000	-	0.5000	-2687.20	5384.41
(0,5)	-	-	-	-40.9847	0.7628	-0.0038	0.0000	0.0000	0.5000	-2687.20	5386.41
(1,2)	-0.0032	-	-	-12.8541	0.1099	-	-	-	0.4917	-2687.75	5383.51
(1,3)	0.0003	-	-	-43.1509	0.8096	-0.0040	-	-	0.4900	2687.15	5384.30
(1,4)	0.0003	-	-	-43.1509	0.8096	-0.0040	0.0000	-	0.4900	-2687.15	5386.30
(2,2)	0.0582	-0.0011	-	-8.7733	0.0710	-	-	-	0.5000	-2686.76	5383.51
(2,3)	0.0652	-0.0012	-	-8.7732	0.0805	-0.0001	-	-	0.5000	-2686.46	5384.92
(3,2)	0.0582	-0.0011	0.0000	-8.7733	0.0710	-	-	-	0.5000	-2686.76	5385.51

Table 3.8: Likelihood ratio test to compare nested Gompertz-Makeham family models for AD and non-AD dementia.

Alternative Model	Nested Model	Degrees of Freedom	p-value
(0,3)	(0,2)	4	0.0204
(1,2)	(0,2)	4	0.03252
(1,3)	(1,2)	5	0.9444
(1,4)	(1,2)	6	0.9766
(2,2)	(1,2)	5	0.8496
(2,3)	(1,2)	6	0.8586
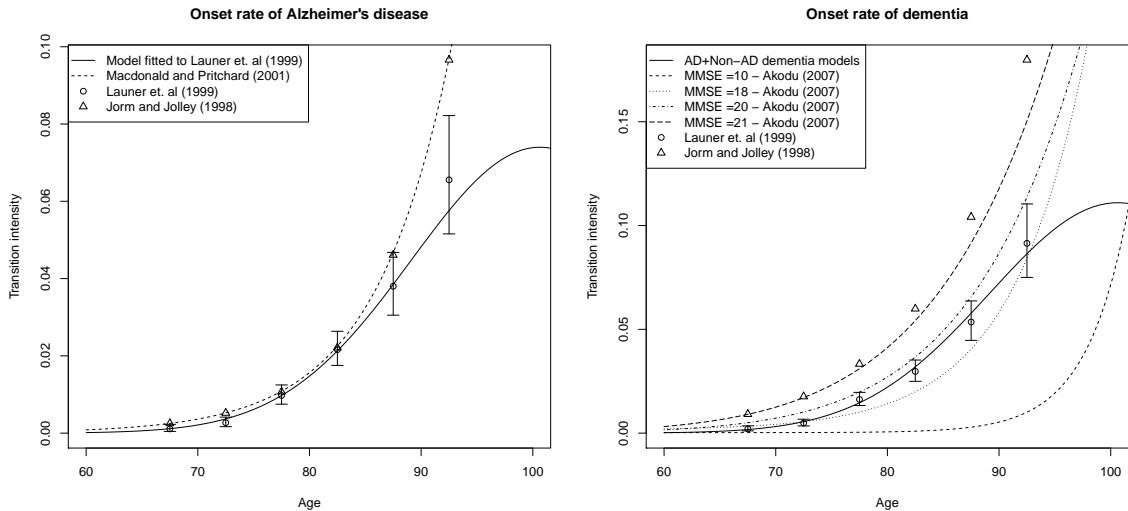
Figure 3.4: Onset of AD and overall dementia (different scales).

For overall dementia we compare to the models fitted by Akodu (2007) to the British CFAS data. The confidence intervals of Launer et al. (1999) are generally between the fitted models for MMSE scores of 18 and 20, perhaps following a pattern for an MMSE score of 19.

As with the AD transition intensities, Jorm and Jolley (1998)'s dementia intensities are significantly greater than those we have used. However, they are very similar to Akodu (2007)'s MMSE score of 21.

Unlike those being compared to, our model has decreasing transition intensities among the oldest old ($>$ 100 years old). There is very little data available for such ages and since any LTC claim would be short (due to the very high mortality at such ages) there would be little distortion of adverse selection costs, we proceed with the transition intensities as they have been fitted.

3.2.7 Post-dementia Mortality

In this section we parameterise the force of mortality for lives with dementia, and lives who have been put into a nursing care facility due to their dementia. We assume the same relative risks due to functional ability as derived in Section 3.2.3 applies after diagnosis of dementia and that the transition intensities fitted here apply to lives with no ADLs.

Several studies have been performed to identify any difference in the survivorship of Alzheimer's type dementia and that of vascular dementia. Burns (1993) summarises the findings of 12 of these studies. While 3 of them (including the largest study) found vascular dementia to have poorer survivorship, the majority could not find any significant difference in the prognosis between the two forms of dementia. There is also a possibility of heterogeneity in the classification: since AD can only be confirmed

through autopsy, there may have been cases of mis-diagnosis of dementia type. Since vascular dementia is the most common of the non-AD dementias, we use this as a proxy for all non-AD types and assume the mortality of lives with any form of dementia as independent of the type of dementia.

Colgan (2006) found the forces of mortality for lives with AD and for those with AD in an institution using data from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD). CERAD was a case-control study of sufferers of AD in the U.S., which aimed to observe the progression of the disease, including the progression into a nursing home, through annual examinations. Confirmation of diagnoses through autopsy was performed where possible. The analysis covers the period 1986–1995, so we assume the rates are effective for a life attaining age x on the 1st of January, 1990.

Macdonald and Pritchard (2000) fitted mortality rates to a very small data set which limited the number of parameters they could fit and chose simple adjustments to their healthy mortality. We use Colgan (2006)'s work since it uses a larger data set and is based on the raw data rather than summary statistics (which Macdonald and Pritchard (2000) used).

The force of mortality after diagnosis but before institutionalisation for males was found to be significantly different to females, so sex specific rates were calculated. For males this was,

$$\mu_x^{10,4} = \exp\left(-8.17 + 0.071x\right), \tag{3.21}$$

and for females it was,

$$\mu_x^{10,4} = 0.029 + \exp\left(-22.014 + 0.221x\right). \tag{3.22}$$

The force of mortality after institutionalisation for males was also found to be significantly different to females, so sex specific rates were calculated. For males it was,

$$\mu_x^{20,4} = \exp\left(-5.13 + 0.052x\right), \tag{3.23}$$

and for females it was,

$$\mu_x^{20,4} = 0.105 + \exp\left(-16.220 + 0.166x\right). \tag{3.24}$$

We apply the same mortality improvements model to each of these, as we used for non-demented mortality (see Section 3.2.4). The results of this are shown in Figures 3.5 and 3.6. For comparison, also shown are the forces of mortality used by Macdonald and Pritchard (2001).

As Colgan (2006) noted, Macdonald and Pritchard (2000)'s pre-institutionalisation are much lower. The results which the latter used, suggested lighter mortality than
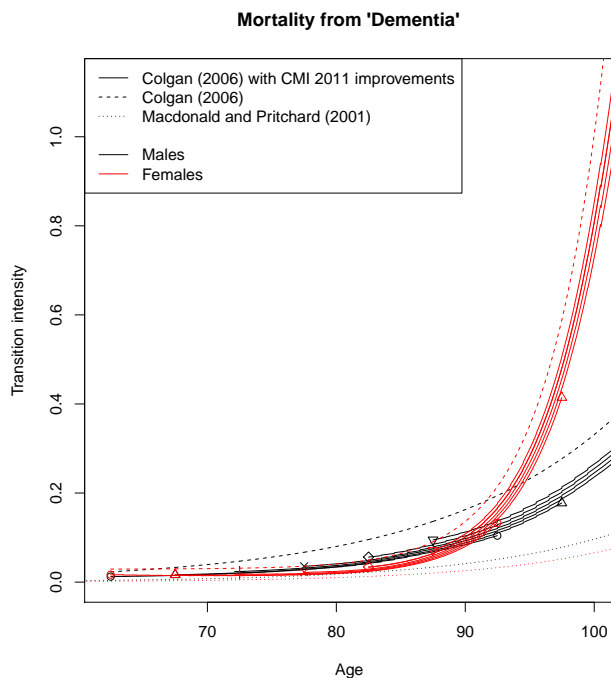
Figure 3.5: Force of mortality from the 'Dementia' state with and without CMI improvements. Improved mortality shown from 2013 with points representing the start and end points of a 30 year period for lives aged $62.5, 67.5, \ldots, 87.5$ in 2013.

that of healthy lives, which they justified as being due to the informal care which sufferers would be receiving while their symptoms progressed.

For the post-institutionalisation mortality, Colgan (2006) is initially lighter and higher at older ages. Colgan (2006) suggested it was due to Macdonald and Pritchard (2000)'s limitation of fitting as a simple function of their baseline mortality, thereby being restricted in the shape.

3.2.8 Institutionalisation After Dementia

In this section we parameterise the progression after being diagnosed with dementia to the state where they require care in an institution.

Ballard et al. (2001) found that the major causes of dementia (Alzheimer's disease, Vascular dementia and Dementia with Lewy Bodies) had similar progression. Moreover, although there have been some links between APOEε4 and the rate of decline in AD, most studies have not found any association (Gauthier et al., 2006). We therefore assume that the cognitive decline is independent of the dementia type, *i.e.* an individual who is diagnosed with dementia will reach a stage where they require nursing care at the same rate regardless of what caused the dementia to occur.

Colgan (2006) fitted a model for transition intensities into an institution after diagnosis with AD to CERAD data. He found no significant difference between males
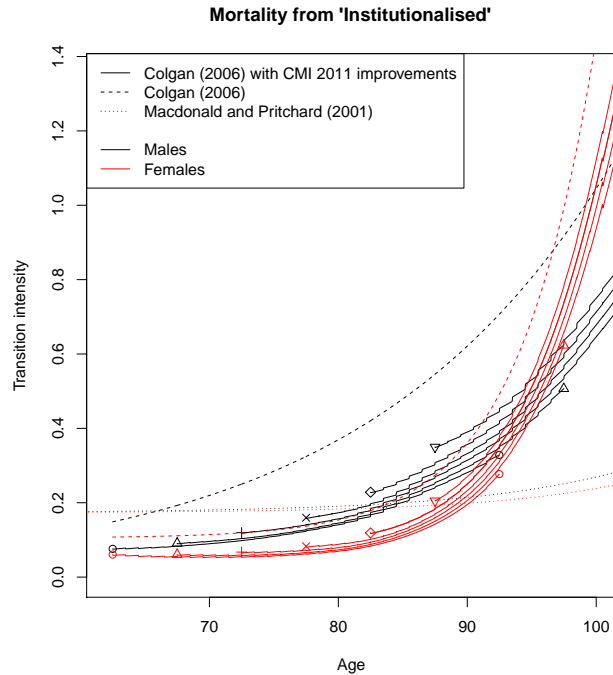
Figure 3.6: Force of mortality from the 'Institutionalised' state with and without CMI improvements. Improved mortality shown from 2013 with points representing the start and end points of a 30 year period for lives aged $62.5, 67.5, \ldots, 87.5$ in 2013.

and females so fitted in aggregate. Under our assumption of equal progression, this is assumed to apply to any dementia, hence,

$$\mu_x^{10,20} = \exp\left(-2.92 + 0.016x\right). \tag{3.25}$$

3.2.9   Undiagnosed Dementia

Since dementia is progressive, it may be noticeable to the life or their family members before a clinical diagnosis. Without details of the disease in medical records, the insurer's ability to detect it in propositions is limited and thereby giving rise to information asymmetry and possible adverse selection in LTC. Introducing to our model (as shown in Figure 3.7), an intermediate stage of cognitive decline where the individual shows some of the signs, *e.g.* memory degrading at an increased rate, will allow us to use a higher rate of insurance purchase from those lives who have noticed this greater need for long term care and assess the cost of this source of adverse selection.

We now need to parameterise the transition intensities between the stages of cognitive function. Choosing the Initial Signs state such that the rate of entry to the state is equal to the rate of exit simplifies the parameterisation to calculating one transition intensity. This is purely a modelling assumption as a means of introducing the possibility of such a state into the model. In reality the rates in and out of this

state could differ — particularly since having suspicions might be expected to cause the life to go to the doctor for diagnosis. Moreover, our Markov assumption may be expected to break down.

By further assuming that mortality is unaffected at this early stage of dementia, all transition intensities are independent of whether the life is healthy or has the initial signs of dementia. This allows us to reduce the parameterisation to models with only the cognitive function transitions without any further loss of generality:

- Model A — consider a 2-state model used to represent a life by whether they have been medically diagnosed. This is shown in Figure 3.8.

- Model B — If we insert a state representing when the life has noticed the initial signs of dementia but before being medically diagnosed, we will get the 3-state model depicted in Figure 3.9.

Let α_t be transition intensity between Healthy and Diagnosed. This can be observed and estimated from prospective studies. However, our intermediate Initial Signs (1) state is at some arbitrary stage between Diagnosis (2) and Healthy (0) and as such, the transition intensities are unavailable. Let β_t be this unknown transition intensity between Healthy and Initial Signs, and between Initial Signs and Diagnosed. In terms of our models which include functional disability, $\mu_{x,g}^{0i,1i} = \alpha_t$, and $\mu_{x,g}^{0,4} = \mu_{x,g}^{4,8} = \mu_{x,g}^{1,5} = \ldots \mu_{x,g}^{7,11} = \beta_t$.

To ensure consistency between both models, we require the probabilities of being diagnosed under models A and B to be equal. Denote by ${}_tA_x^{ij}$, the probability of being in state j at age $x + t$ conditional on being in state i at age x under model A and similarly for model B, ${}_tB_x^{ij}$. If we assume no lives have developed the initial signs at age x, our equivalence requirement can be represented by ${}_tA_x^{02} = {}_tB_x^{02}$ or alternatively,

$$ {}_tA_x^{00} = {}_tB_x^{00} + {}_tB_x^{01}. \tag{3.26} $$

Clearly ${}_tB_x^{00} = \exp\left(-\int_0^t \beta_{x+s}\,ds\right)$ and similarly ${}_tA_x^{00} = \exp\left(-\int_0^t \alpha_{x+s}\,ds\right)$. For ${}_tB_x^{01}$ we have,

$$
\begin{aligned}
{}_tB_x^{01} &= \int_0^t {}_sB_x^{00}\,\beta_{x+s}\,{}_{t-s}B_{x+s}^{11}\,ds \\[2ex]
&= \int_0^t \exp\left(-\int_0^s \beta_{x+r}\,dr - \int_s^t \beta_{x+r}\,dr\right)\beta_{x+s}\,ds \\[2ex]
&= \exp\left(-\int_0^t \beta_{x+r}\,dr\right)\int_0^t \beta_{x+s}\,ds \\[2ex]
&= {}_tB_x^{00}\int_0^t \beta_{x+s}\,ds.
\end{aligned}
\tag{3.27}
$$
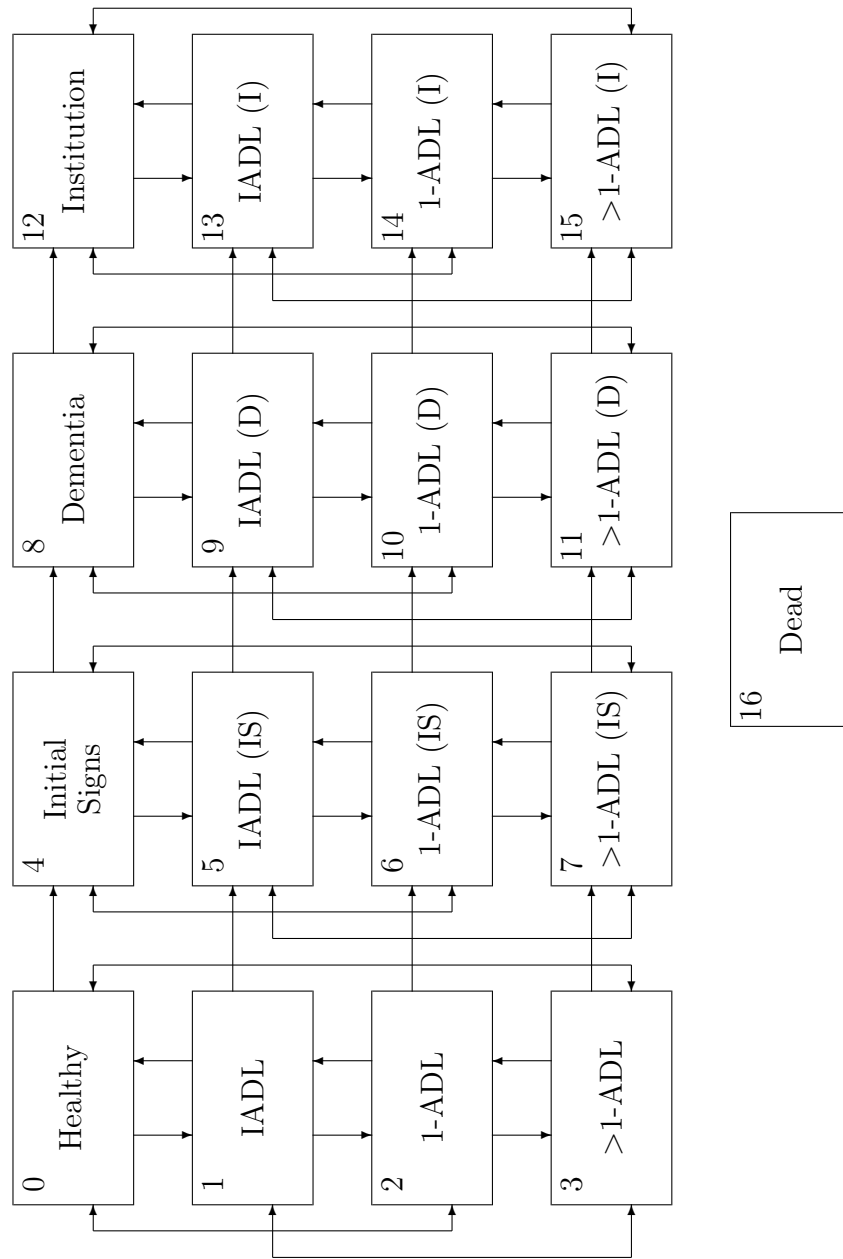
Figure 3.7: A Markov model of functional ability and cognitive function with a stage in cognitive decline where the initial signs of dementia have not been diagnosed but are visible to the individual. The arrows to the Dead state are omitted but may be entered from any state.
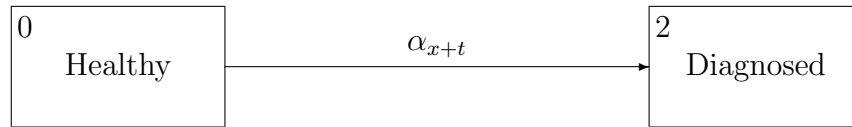
Figure 3.8: Model A: a 2-state model of the onset of dementia

Figure 3.9: Model B: a 3-state model of the onset of dementia

Inserting Equation (3.27) into Equation (3.26) yields a differential equation for β_t in terms of α_t:

$$\exp\left(-\int_0^t \alpha_{x+s}\, ds\right) = \exp\left(-\int_0^t \beta_{x+s}\, ds\right)\left(1 + \int_0^t \beta_{x+s}\, ds\right)$$

$$\exp\left(-\int_0^t (\alpha_{x+s} - \beta_{x+s})\, ds\right) = 1 + \int_0^t \beta_{x+s}\, ds$$

$$-\int_0^t (\alpha_{x+s} - \beta_{x+s})\, ds = \log\left(1 + \int_0^t \beta_{x+s}\, ds\right)$$

$$\beta_{x+t} - \alpha_{x+t} = \frac{d}{dt}\log\left(1 + \int_0^t \beta_{x+s}\, ds\right)$$

$$= \frac{\beta_{x+t}}{1 + \int_0^t \beta_{x+s}\, ds}$$

$$\alpha_{x+t} = \beta_{x+t}\left(1 - \frac{1}{1 + \int_0^t \beta_{x+s}\, ds}\right)$$

$$\beta_{x+t} = \alpha_{x+t}\left(\frac{1 + \int_0^t \beta_{x+s}\, ds}{\int_0^t \beta_{x+s}\, ds}\right). \tag{3.28}$$

To avoid β_{x+t} exploding to infinity at $t = 0$, for some small time step h, we assume $\beta_x = \beta_{x+s}$, for $s \in [0, h]$. This assumption allows us to calculate β_x as

$$\beta_x = \beta_{x+h} = \alpha_{x+h}\left(\frac{1 + \beta_{x+h}h}{\beta_{x+h}h}\right)$$

$$= \frac{\alpha_{x+h}h \pm \sqrt{\alpha_{x+h}^2 h^2 + 4\alpha_{x+h}h}}{2h}. \tag{3.29}$$

3.3 Simulating Long-Term Care Insurance Payments

In this section we describe how simulation can be used to estimate the expected present values of benefits and premiums in a long-term care insurance contract.

3.3.1 Notation

We first define the notation introduced in this section which will be used in later sections.

1. The force of interest and force of inflation are assumed constant at rates δ *per annum* and ν *per annum* respectively.

2. The set of periods which a policy may go through is denoted by $\mathcal{P}$, and consists of P — after inception but before any claim; C — after a claim on a functional disability but before dementia; D — after the first claim involving dementia.

For a life aged x at the simulation start date, of sex ς, with genotype g and insured at time t, from health state j:

3. The smoothed probability that this life is in period ρ at the start of calendar year c is denoted by ${}_c p_{x,\varsigma,g,t}^{j,\rho}$.

4. Given this life is in period ρ, the smoothed average present value of future benefits and premium income are denoted by $A_{x,\varsigma,g,t:c}^{j,\rho}$ and $a_{x,\varsigma,g,t:c}^{j,\rho}$, respectively.

3.3.2 Reason for Simulation Approach

The model of health depicted in Figure 3.7 is Markov — transition intensities depend on the state currently occupied, not the path to get there. However, LTC payments could be capped by the insurer with a limit on annual payouts or some total sum assured, or the government could limit the liability faced by the individual before state support is paid (a policy currently under discussion in the U.K. which we discuss further in Chapter 4). In cases such as these, the cashflow at any point in time depends not only on the current health status, but on the entire history of health statuses.

This disjoint nature between the cashflows and current health status creates a difficulty in calculating the expected present value of benefits for the purpose of setting premiums. Consider a contract with all the limits as described above, bought by a life aged x, when in state i:

- The insurer limits total of payments over the lifetime of the policy to sum assured, s and the benefits payable within a calendar year to, a.

- The government pays for care costs once the individual's care liability (which may be paid by the insurer) has surpassed, g.

- 'Hotel' costs (accommodation cost in a care home) are excluded from the government cap but covered by the insurer, up to its own benefit limits.

- The care and hotel costs for state j, amount to c_j and h_j, respectively.

- All costs and caps increase over time at the same constant force of inflation, denoted by ν. All benefit payments and caps are compared at the same purchasing power.

Let $_tI_x^j$ be an indicator random variable such that,

$$
_tI_x^j = \begin{cases} 1 & \text{if } x, \text{ is in } j \text{ at age } x+t \\ 0 & \text{otherwise.} \end{cases} \tag{3.30}
$$

Let $I(f)$ be another indicator function such that,

$$
I(f) = \begin{cases} 1 & \text{if } f > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.31}
$$

We denote the random variable representing the care cost if x is in state j, at time t, as a continuously payable annual rate, by $_tC_x^j$ and calculate it as,

$$
_tC_x^j = c_j e^{\nu t} I\left(g - \sum_l \int_{-x}^t {}_rI_x^l {}_rC_x^l e^{-\nu r}\, dr \right). \tag{3.32}
$$

We further denote the random variable representing the insurance benefit payment if x is in state j, at time t, as a continuously payable annual rate, by $_tB_x^j$, and calculate it as,

$$
_tB_x^j = \left({}_tC_x^j + h_j e^{\nu t} \right) \quad I\left(s - \sum_l \int_0^t {}_rI_x^l {}_rB_x^l e^{-\nu r}\, dr \right)
$$

$$
\times I\left(a - \sum_l \int_{\lfloor t \rfloor}^t {}_rI_x^l {}_rB_x^l e^{-\nu r}\, dr \right). \tag{3.33}
$$

The expected present value of benefits paid to x can then be calculated as,

$$
E(\text{PV Benefits payable to } x) = E\left(\sum_j \int_0^\infty {}_rB_x^j {}_rI_x^j e^{-\delta r}\, dr \right). \tag{3.34}
$$

Equation (3.34) involves a complicated system of integration which is simplest to solve approximately using simulation. We perform simulations of future lifetimes and

the consequent LTC benefit cashflows that occur (by solving Equations 3.32 and 3.33). Equation (3.34) can then be found by taking the average of the discounted values of our simulated benefit cashflows. If we assume that no care liability has been faced prior to purchasing insurance, we avoid the complication of setting premium rates for a varying degree of potential benefit sizes (as limited by the government cap) and simplify our pricing problem to simulating from the point of insurance purchase.

3.3.3 Simulation Method

To simulate life histories, represented by transitions between states in our health model, we first find the time of a transition. By conditioning on the transition occurring at this time, we then find the state the life moves to. Denote the random variables representing the time of the ith transition by, T_i, and S_i, as the state a life is in after the ith transition. Further, denote the value taken by them as, t_i and s_i, respectively. We assume a life aged, x, at the start of the simulation is in state, s_0, at time, $t_0 = t$. For a life in state, s_{i-1}, at time, t_{i-1}, we simulate the next transition time using the inverse transform method (Devroye, 1986) solving $F_{T_i}(t_i) = u$ where u is an observation from the Uniform $[0,1]$ distribution and

$$F_{T_i}(t_i) = 1 - \exp\left(-\int_{t_{i-1}}^{t_i} \sum_{j \in \mathcal{S}} \mu_{x+u}^{s_{i-1},j} du\right), \tag{3.35}$$

for t_i. These integrals are solved using Simpson's rule with a step size of 2^{-11} and our simulation period is 30 years to capture most of a life's future lifetime.

Given a life makes the ith transition at time t_i, the probability that the movement was to state j is

$$P(S_i = s_i | T_i = t_i) = \frac{\mu_{x+t_i}^{s_{i-1},s_i}}{\sum_{j \in \mathcal{S}} \mu_{x+t_i}^{s_{i-1},j}}. \tag{3.36}$$

Define the conditional cumulative distribution function as

$$F_{S_i | T_i = t_i}(s_i) = P(S_i \leq s_i | T_i = t_i) = \frac{\sum_{j=0}^{s_i} \mu_{x+t_i}^{s_{i-1},j}}{\sum_{j \in \mathcal{S}} \mu_{x+t_i}^{s_{i-1},j}}. \tag{3.37}$$

We simulate the jump at time t_i by applying the inverse transform method on $F_{S_i | T_i = t_i}(s_i) = v$, where v is an observation from the Uniform $[0,1]$ distribution, solving for s_i.

With knowledge of the future lifetime of our simulated life, we can calculate the present value payments in respect of a long-term care insurance contract bought at time t_0. By averaging over a large number of simulations, we calculate an estimate of

Equation (3.34). To allow a product design with regular premiums, we also estimate the expected present value of an annuity in premium paying states, which we refer to as the premium annuity. The premium annuity and benefit payments are increased at the same force of inflation, ν *per annum*. As in our above example, all benefit payments and caps are compared at the same purchasing power.

A deferred period can be set, such that benefits do not commence payment until the life has been in benefit paying states for longer than the deferred period. Associated with the deferred period, we can define a linked claim period, if a life recovers after the benefit payment has commenced and subsequently enters a claim state within the linked claim period, the benefit payment will recommence immediately instead of having to wait until the end of the deferred period. This is a common feature of products with deferred periods *e.g.* income protection insurance, to allow for relapses in illness, however the Markov nature of our model would prevent any true relapse.

We assume that a life cannot buy insurance if it would put them into a claim immediately. Additionally, since we are primarily concerned with the potential cost of adverse selection, we do not model the purchase of insurance after dementia has been diagnosed — this is observable to an insurer so we assume they can underwrite them appropriately. We call the set of remaining states insurable.

Since we simulate from the point of insurance purchase, we require the EPVs from each time step at which insurance can be purchased, which we refer to as purchase time steps. Due to the computing time required, we restrict our purchase time steps to the 30 year period in steps of 0.015625 years. Simulations are performed for 50,000 lives at each purchase time step, for males and females of each APOE genotype for each insurable state. Thus we perform $50,000 \times 2 \times 5 \times 6 = 3,000,000$ simulations at each purchase time step.

3.3.4   Information Gained by the Insurer

Integral to adverse selection is the notion of asymmetric information — where the policyholder knows more about their health than the insurer. Over time, the insurer can learn about the mix of business it has on its books, based on its claims experience. By better understanding the mix of lives to whom they have been selling, the insurer will be better placed to estimate the mix of future business and the premiums to charge. This will form the basis for the dynamic premium rates of Section 3.4.2.

From details of claims payouts, an insurer could observe 3 key events over the lifetime of a policy, which will change their knowledge of the mix of business:

- Inception — At the point of inception, the insurer has no information on the actual mix of business beyond what is implied in the pricing basis.

- A claim for functional disability (unrelated to dementia) — The insurer now

knows what functional disability state the life is in; transition between states of functional disability is determined only by current functional disability.

- Claim involving dementia — When the insurer knows the life has dementia (we assume they also learn of functional disability at this point also), everything relevant to its future transitions is now known and no more useful information is gleaned from further claims.

These observations will tell them more about the mix of the policyholders than was known at underwriting and allow them to more accurately value future cashflows in the periods following them:

- P — After inception but before any claim.

- C — After the first functional disability claim but before a claim involving dementia.

- D — After the first claim involving dementia.

The set of these periods can therefore be written as, $\mathcal{P} = \{P, C, D\}$. We explain further by use of an example: Assume some benefit is payable to an insured life while in states $\{3, 7, 10, 11, 12, 13, 14, 15\}$. Consider life x, who bought his insurance from state 5 at time t_0. The insurer cannot distinguish him from any other life in states $\{0, 1, 2, 4, 5, 6\}$.

1. x reaches state 7 at time t_1, making a claim unrelated to dementia. The insurer now knows he is in either state 3 or state 7 and can improve its modelling in respect of this.

2. x recovers to state 6 at time t_2.

3. x subsequently claims again in state 7 at time t_3. In a large portfolio of business, this does not tell the insurer anything new because x's transitions between functional disability are being modelled appropriately already based on being in either state 3 or state 7 at time t_1 and the Markov nature of the model (recall that functional ability transitions are assumed to be independent of cognitive function).

4. x develops dementia and claims from state 11 at time t_4. The insurer knows x is in state 11 and updates its model with this information.

5. x's claim ceases upon movement to state 8 at time t_5.

6. x dies at time t_6.

In this example, period P was from t_0 until t_1; period C was from t_1 until t_4; and period D was from t_4 until the policy ended at t_6.

To model this information development, at the start of each calendar year of our simulations, we calculate the probability that a life, who purchased insurance at time t, is in each period. Additionally, we estimate the expected present value of all future benefits and of future income (including those cashflows which will occur within a subsequent period) for lives who are in each of our periods.

3.3.5   Smoothing

The simulated expected present values and probabilities within each calendar year c are smoothed with respect to insurance purchase time $t \leq c + 1$, by fitting cubic splines using the ALGLIB package for C++ (Bochkanov and Bystritsky, 2013). This removes the random noise and gives us a means of estimating the expected present values, $A^{j,\rho}_{x,\varsigma,g,t:c}$ and $a^{j,\rho}_{x,\varsigma,g,t:c}$, and the probability of being in period ρ, $_cp^{j,\rho}_{x,\varsigma,g,t}$, for any insurance purchase date $t \leq c + 1$. Splines are fitted to these independently for every combination of initial age, sex, genotype, policy period and the health state which insurance was purchased from.

To explain further what smoothing is performed, in Table 3.9 we present the simulated present values of benefits for males in period P, with genotype $\varepsilon 2\varepsilon 2$, who were aged 62.5 at calendar year 0 and buy insurance at time t from state 0. In each calendar year for which the life has not yet made any claim, all future premium payments are discounted to the start of the year. Purchases at integer times are considered to be in force in the year before to provide a knot for fitting the final section of the year — in the table, when $t = 1$, there is a value for $c = 0$. Smoothing is then performed on each column independently, with respect to t. The results of smoothing calendar year 3 are shown in Figure 3.10.

3.4   Long-Term Care Insurance Market

In this section we set up a model for the market of LTC by adding states our health model of Figure 3.7, to represent the presence of insurance. To allow us to set higher (or lower) rates of insurance purchase from lives who have received a genetic test (and hence allow this as a source of adverse selection), we also add states representing whether a genetic test has been received. This market model is depicted in Figure 3.11.

We name each state by reference to its insured status, its test status, $\iota \in \mathbb{B}$, and

Table 3.9: Simulated present values of future benefits for males in period P, with genotype $\varepsilon2\varepsilon2$, aged 62.5 at calendar year 0, buying insurance from state 0.

Purchase	Calendar year (c)		
Time (t)	0	1	2
0.000000	6644.56	6608.76	6458.43
0.015625	6638.55	6579.41	6421.86
0.031250	6616.18	6575.82	6437.73
⋮			
0.953125	6370.83	6431.86	6360.77
0.968750	6302.85	6364.04	6296.52
0.984375	6446.14	6507.69	6436.21
1.000000	6378.06	6441.84	6403.86
1.015625	-	6354.15	6300.15
1.031250	-	6468.30	6418.69
⋮			
1.968750	-	6114.98	6174.80
1.984375	-	6224.97	6287.52
2.000000	-	6134.46	6195.80
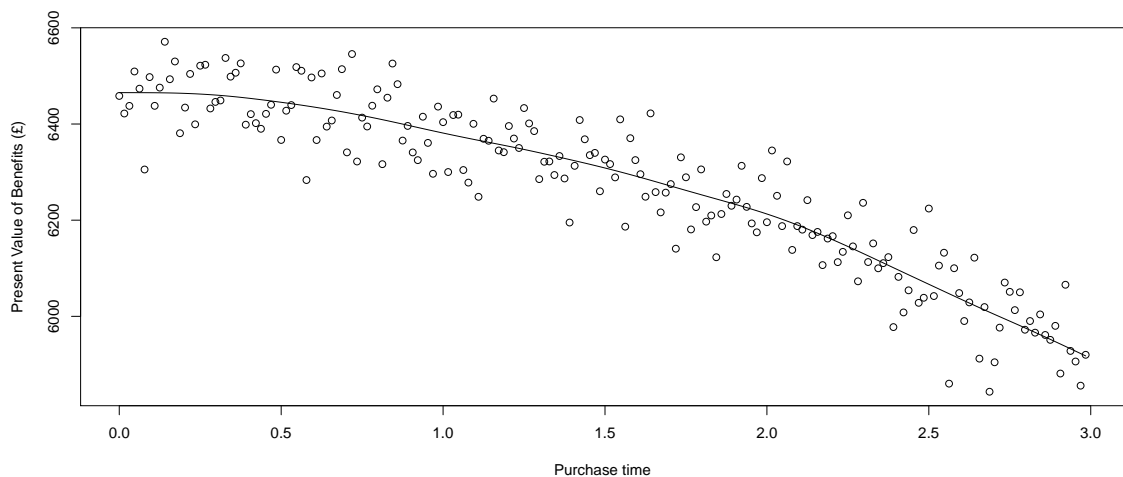
Figure 3.10: Simulated and smoothed present values of future benefits, as at the start of calendar year 2, for males in period P, with genotype $\varepsilon2\varepsilon2$, who were aged 62.5 at calendar year 0, buying insurance from state 0.

its health status, $i \in \mathcal{S}$, as $\iota\vartheta i$, where $\mathbb{B} = \{0, 1\}$,

$$\vartheta = \begin{cases} 0 & \text{if the life is untested} \\ 1 & \text{if the life is tested,} \end{cases} \tag{3.38}$$

and similarly,

$$\iota = \begin{cases} 0 & \text{if the life is uninsured} \\ 1 & \text{if the life is insured.} \end{cases} \tag{3.39}$$

3.4.1 Notation

We introduce notation for this section:

1. The set of states belonging to underwriting class k is denoted by $\mathcal{U}_k$.

2. Let $p^i_{x,\varsigma,g}$ be the probability at the start of modelling, that a life aged x is of sex ς, has genotype g, and is in state i. This will be specific to the market to which the model is being applied so we calculate it in Chapter 4 (Section 4.2.2), where we apply our modelling to the U.K. market, but set out how it will be used in formulae in this section.

3. The conditional probability that a life aged $x + t$ who purchased insurance into class k, at time t, was of sex ς, with genotype g, and insured from health state j, is denoted by $\eta^{j|k}_{\varsigma,g|x,t}$.

Consider a life aged x at the start of our modelling (1st January, 2013), of sex ς, with genotype g.

3. The probability this life is in state def at age $x + t$, given they were in state abc at age x, is denoted by ${}_t p^{abc,def}_{x,\varsigma,g}$.

4. $\mu^{abc,def}_{x,t,\varsigma,g}$ is the transition intensity between state abc and state def, for this life at time t.

5. Let $\varpi^k_{x,t}$ and $\omega^k_{x,t}$ denote the regular (paid annually) and single (initial lump sum) premium respectively, that would be charged to a life aged $x + t$ at time t, insured into underwriting class k, if there was no adverse selection. We will refer to this as the base premium.

6. Let $\overline{\psi}^k_{x,c}$ and $\psi^k_{x,c}$ denote the repricing adjustment made in calendar year c, to the underwriting class k base premium for regular premium and single premium contracts respectively.

7. The premium charged as a result of repricing activity is denoted by, $\overline{\Pi}^k_{x,t}$ and $\Pi^k_{x,t}$, for regular and single premium respectively.
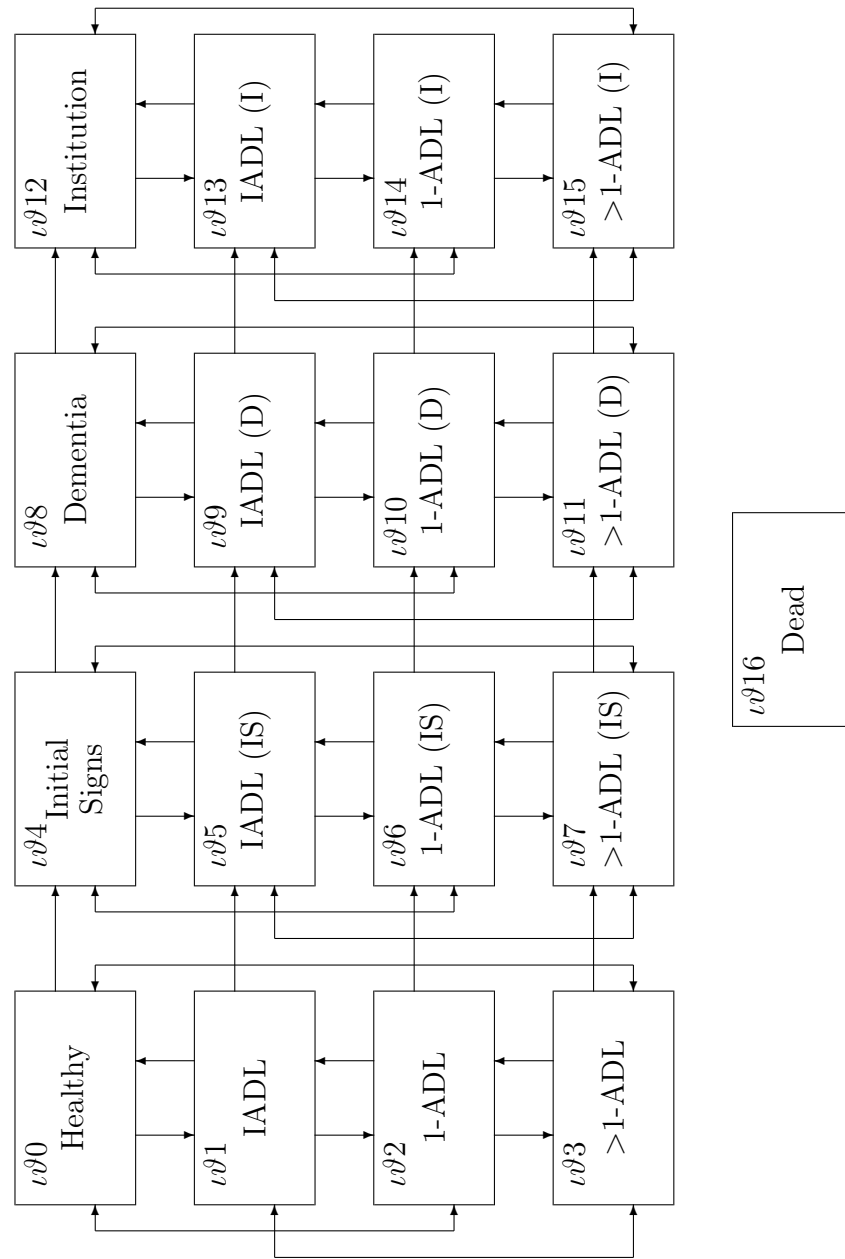
Figure 3.11: A Markov model of an insurance market for LTC where ι denotes whether the life is insured, and ϑ denotes whether a genetic test has been performed. The arrows to the Dead state are omitted but may be entered from any state.

9. Let $B_{x,t:c}^{k,\rho}$ and $I_{x,t:c}^{k,\rho}$ be stochastic processes representing the present value of future benefit payments and premium income respectively as at the start of calendar year c, for a life who bought insurance in underwriting class k, at time t and is now in period ρ. Their associated filtration is denoted $\mathcal{F}_c$ at calendar year c.

To distinguish between the pricing basis and the experience basis, let the addition of a tilde to the equivalent pricing basis notation, denote experience basis.

3.4.2   Dynamic Premium Rates

Our prime interest is in how the costs of adverse selection will develop over time after setting up a new market for insurance. This will be related to information gained by the insurer, from its claims experience, discussed in Section 3.3.4: In a start-up market, the insurer has no experience with which to properly set a pricing basis. They would make assumptions as to their business mix, but this might not be borne out in their sales if the product appealed more to one group and less to another. Where the differences between groups are noticeable, underwriters could assign additional premiums to high risk groups and still have *actuarially fair* premiums for all. However, where the details of an individual's risk is hidden from the insurer at policy inception, the insurer would not learn of the difference from its assumed mix until some later time, when it observes a higher than expected volume of claims. As the insurer learns who is buying the product, they can reassess their assumptions of business mix and reprice appropriately. In doing so, the can reduce the adverse selection costs which they face.

Our aim is to model this repricing in such a way as to reflect how an insurance company would behave. We assume it occurs annually and is not to recover past losses, but to try to set premiums sufficient to cover the benefits of the lives who are actually buying the product at that time (recovering past losses might be deemed to be treating customers unfairly). Additionally, we assume that they treat cohorts independently. In this way, the pricing of a group of policies that is split into age groups or underwriting classes, is influenced by its own experience but not that of any other age group or underwriting class.

We will consider two versions of insurance contract: single premium and regular premium paid continuously while the policyholder is not being paid benefits. Single premiums can be large sums of money, particularly for a life living on a pension. Regular premium could be more marketable if it is affordable. The potential costs of adverse selection in a regular premium design are larger than in a single premium — benefit costs in each will be the same however premium income will be lower than assumed since it is not paid upfront. Formulae will be presented for unisex premiums,

in accordance with European Court of Justice (2011), but we will also consider a version with gender specific premiums to analyse the cross-subsidy between the sexes. In that case, we can calculate sex-specific equivalents for all formulae given below by setting the purchase rate for the opposite sex to zero. Where unisex premiums are charged, since the Court's ruling does not explicitly prohibit asking a life's sex, we assume the company asks this on the application form. The unisex premium rates charged at time, t, therefore reflect the actual mix of sex to which business has been sold.

The premiums for an underwriting class will be calculated using the equivalence principle, whereby the expected present values of outgo and of premium income within an underwriting class are equal. The pricing basis will assume there is no adverse selection. We do not include any expenses, nor are there any loadings for profit, so outgo is limited to benefit payments.

These expectations are conditional on a life purchasing insurance at time t. The conditional probability that a life who purchased insurance at time t was of sex ς, with genotype g and insured from health state j, $\eta^{j|k}_{\varsigma,g|x,t}$, is calculated as,

$$
\eta^{j|k}_{\varsigma,g|x,t} = \frac{\displaystyle\sum_{i\in\mathcal{S},\vartheta\in\mathbb{B}} p^i_{x,\varsigma,g}\, {}_tp^{00i,0\vartheta j}_{x,\varsigma,g}\, \mu^{0\vartheta j,1\vartheta j}_{x,t,\varsigma,g}}{\displaystyle\sum_{i\in\mathcal{S},\vartheta\in\mathbb{B},l\in\mathcal{U}_k,g\in\mathcal{G}} \left(p^i_{x,M,g}\, {}_tp^{00i,0\vartheta l}_{x,M,g}\, \mu^{0\vartheta l,1\vartheta l}_{x,t,M,g} + p^i_{x,F,g}\, {}_tp^{0\vartheta l,1\vartheta l}_{x,F,g}\, \mu^{0\vartheta l,1\vartheta l}_{x,t,F,g} \right)}. \tag{3.40}
$$

The probabilities ${}_tp^{00i,0\vartheta j}_{x,\varsigma,g}$ are found by solving the Kolmogorov forward equations:

$$
\frac{d}{dt}{}_tp^{00i,0\vartheta j}_{x,\varsigma,g} = \sum_{(k,l)\neq(\vartheta,j)} \left({}_tp^{00i,0kl}_{x,\varsigma,g}\, \mu^{0kl,0\vartheta j}_{x,t,\varsigma,g} - {}_tp^{00i,0\vartheta j}_{x,\varsigma,g}\, \mu^{0\vartheta j,0kl}_{x,t,\varsigma,g} \right) - {}_tp^{00i,0\vartheta j}_{x,\varsigma,g}\, \mu^{0\vartheta j,1\vartheta j}_{x,t,\varsigma,g}. \tag{3.41}
$$

This was done using a 4th order Runge Kutta algorithm with a step size of 2^{-12} and boundary conditions ${}_0p^{00i,jkl}_{x,\varsigma,g} = \delta_{00i,jkl}$, where $\delta_{00i,jkl}$ is the Kronecker delta.

The unisex premium needs to take account of the actual sex mix since this is known at outset, so we scale the respective expected present values to the size of business from each sex. The unisex base regular premium for a life aged x at outset and insured into class k at time t (in calendar year $c = \lfloor t \rfloor$), is calculated as

$$
\varpi^k_{x,t} = \frac{\displaystyle\sum_{\varsigma\in\{M,F\}} \left[\frac{\displaystyle\sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \tilde{\eta}^{j|k}_{\varsigma,g|x,t}}{\displaystyle\sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta^{j|k}_{\varsigma,g|x,t}} \sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta^{j|k}_{\varsigma,g|x,t} A^{j,P}_{x,\varsigma,g,t:c} \right]}{\displaystyle\sum_{\varsigma\in\{M,F\}} \left[\frac{\displaystyle\sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \tilde{\eta}^{j|k}_{\varsigma,g|x,t}}{\displaystyle\sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta^{j|k}_{\varsigma,g|x,t}} \sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta^{j|k}_{\varsigma,g|x,t} a^{j,P}_{x,\varsigma,g,t:c} \right]}. \tag{3.42}
$$

In the single premium version, the base premium for a life aged x entering an

insured state in $\mathcal{U}_k$ at time t, is calculated as

$$\omega_{x,t}^k = \sum_{\varsigma \in \{M,F\}} \left[\frac{\sum\limits_{j \in \mathcal{U}_k, g \in \mathcal{G}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}}{\sum\limits_{j \in \mathcal{U}_k, g \in \mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}} \sum_{j \in \mathcal{U}_k, g \in \mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k} A_{x,\varsigma,g,t:c}^{j,P} \right]. \tag{3.43}$$

The premiums actually charged to policyholders will reflect the insurer's acquired knowledge of the mix of lives who have been buying it. The method we use to calculate the premiums to charge does not change the underlying assumptions in the pricing basis. Instead, we find a repricing adjustment factor, $\overline{\psi}_{x,\lfloor t \rfloor}^k$, by which to multiply the base premiums to represent the repricing we have assumed occurs annually. The regular premium rate charged to lives aged x at the start of our modelling who buy insurance into class k at time t, is calculated as

$$\overline{\Pi}_{x,t}^k = \overline{\psi}_{x,\lfloor t \rfloor}^k \varpi_{x,t}^k, \tag{3.44}$$

and similarly for single premium, $\Pi_{x,t}^k$.

We calculate the repricing adjustment by comparing the premium rates charged to existing customers, with what would be charged to these same customers given what is now known of their experience, to find how much larger their premium should have been. This can be expressed as

$$\text{Premium adjustment} = \frac{\text{Actual Benefit Payments} + \text{Expected Future Benefits}}{\text{Actual Premium Income} + \text{Expected Future Income}}. \tag{3.45}$$

The insurance company would be able to observe the actual benefit payments and premium income up to the point of analysis. Their expectations for the future will reflect the knowledge of their business mix which they have gained from their claims experience. In our model, we estimate them by considering the present value of future benefit and income as stochastic processes, $B_{x,t:c}^{k,\rho}$ and $I_{x,t:c}^{k,\rho}$, respectively. Note that for single premium,

$$I_{x,t:c}^{k,\rho} = \begin{cases} \Pi_{x,t}^k & \text{if } c = \lfloor t \rfloor \text{ and } \rho = P \\ 0 & \text{if } c > \lfloor t \rfloor \text{ or } \rho \neq P. \end{cases} \tag{3.46}$$

The basis for calculation of the expected value of these stochastic processes will reflect the information gained from the insurance pool's claims history. Hence the claims history up to year c forms filtration, $\mathcal{F}_c$. We rewrite the components of Equation

(3.45) for business sold at time t, in terms of our stochastic processes as,

$$E\left(B_{x,t:\lfloor t\rfloor}^{k,P}|\mathcal{F}_c\right) = \text{Actual Benefit Payments} + \text{Expected Future Benefits}$$

$$= E\left(B_{x,t:\lfloor t\rfloor}^{k,P} - \sum_{\rho\in\mathcal{P}}B_{x,t:c}^{k,\rho}|\mathcal{F}_c\right) + \sum_{\rho\in\mathcal{P}}E\left(B_{x,t:c}^{k,\rho}|\mathcal{F}_c\right) \qquad (3.47)$$

$$E\left(I_{x,t:\lfloor t\rfloor}^{k,P}|\mathcal{F}_c\right) = \text{Actual Premium Income} + \text{Expected Future Income}$$

$$= E\left(I_{x,t:\lfloor t\rfloor}^{k,P} - \sum_{\rho\in\mathcal{P}}I_{x,t:c}^{k,\rho}|\mathcal{F}_c\right) + \sum_{\rho\in\mathcal{P}}E\left(I_{x,t:c}^{k,\rho}|\mathcal{F}_c\right) \qquad (3.48)$$

We calculate Equations 3.47 and 3.48 by calculating each part in turn, first 'Actual' and secondly 'Expected Future' using our simulations.

To calculate the 'Actual' component, for payments up to calendar year c, we calculate their expected values on the experience basis. For benefits this is calculated as,

$$E\left(B_{x,t:\lfloor t\rfloor}^{k,P} - \sum_{\rho\in\mathcal{P}}B_{x,t:c}^{k,\rho}|\mathcal{F}_c\right)$$

$$= \sum_{j\in\mathcal{U}_k,g\in\mathcal{G},\varsigma\in\{M,F\}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}\left(A_{x,\varsigma,g,t:\lfloor t\rfloor}^{j,\rho} - \sum_{\rho\in\mathcal{P}}{}_cp_{x,\varsigma,g,t}^{j,\rho}A_{x,\varsigma,g,t:c}^{j,\rho}e^{(c-\lfloor t\rfloor)(\nu-\delta)}\right),$$

$$(3.49)$$

Similarly, for regular premium income this is calculated as

$$E\left(I_{x,t:\lfloor t\rfloor}^{k,P} - \sum_{\rho\in\mathcal{P}}I_{x,t:c}^{k,\rho}|\mathcal{F}_c\right)$$

$$= \overline{\Pi}_{x,t}^k \sum_{j\in\mathcal{U}_k,g\in\mathcal{G},\varsigma\in\{M,F\}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}\left(a_{x,\varsigma,g,t:\lfloor t\rfloor}^{j,\rho} - \sum_{\rho\in\mathcal{P}}{}_cp_{x,\varsigma,g,t}^{j,\rho}a_{x,\varsigma,g,t:c}^{j,\rho}e^{(c-\lfloor t\rfloor)(\nu-\delta)}\right).$$

$$(3.50)$$

Next we calculate the expected future benefit payments and regular premium income on policies in each of the periods in $\mathcal{P}$. We assume the insurer cannot discern between lives within an underwriting class beyond gender, *i.e.* when a policy is sold, they do not know the particular functional disability status, cognitive ability type or genotype. Each piece of information hidden from the insurer (genotype, functional ability and cognitive function) relates either to the onset of dementia or functional disability but not both. At the point of a claim, the insurer will gain information and

can update their assumptions over the mix of business in the pool.

When a life buys insurance at time t, and until the first claim, the insurer sees it as any other within the underwriting class it is written. Hence the expected benefits and premium income from lives in the P period are calculated on the pricing basis, scaled to the volume of business remaining in this pool. For the benefits, this can be expressed as

$$
E\left(B_{x,t:c}^{k,P}|\mathcal{F}_c\right) = \sum_{\varsigma\in\{M,F\}} \left[\frac{\sum\limits_{j\in\mathcal{U}_k,g\in\mathcal{G}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P}}{\sum\limits_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P}} \sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P} A_{x,\varsigma,g,t:c}^{j,P} \right],
$$
(3.51)

and for similarly for regular premium income

$$
E\left(I_{x,t:c}^{k,P}|\mathcal{F}_c\right) = \overline{\Pi}_{x,t}^{k} \sum_{\varsigma\in\{M,F\}} \left[\frac{\sum\limits_{j\in\mathcal{U}_k,g\in\mathcal{G}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P}}{\sum\limits_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P}} \sum_{j\in\mathcal{U}_k,g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,P} a_{x,\varsigma,g,t:c}^{j,P} \right].
$$
(3.52)

At the point of the first claim unrelated to dementia, because no benefit payments have yet been made and the movement around health states is Markov, the initial functional disability state becomes unimportant since the insurer knows what functional state the life is in at the time of claim. However, because lives move between the functional disability states at the same rate regardless of cognitive function, a claim unrelated to dementia does not tell the insurer anything about cognitive ability type so the insurer must assume the mix to be as in their pricing assumptions. Denote the set of states that belong to the same functional disability type as state i which would be underwritten to class k as $\mathcal{D}_{i,k}$. We calculate the expected future benefit payments and premium income for each functional disability type at entry on the expected basis and scale this for the volume of business which is in the C period (*i.e* the volume of business which has claimed for functional disability but not dementia). Let $V_{x,\varsigma,t:c}^{i,C} = \sum\limits_{j\in\mathcal{D}_{i,k},g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,C}$ and $\tilde{V}_{x,\varsigma,t:c}^{i,C} = \sum\limits_{j\in\mathcal{D}_{i,k},g\in\mathcal{G}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,C}$ be the volume of business at the start of calendar year c of lives aged x in 2013 and insured from the same functional disability type as state i at time t on the pricing and experience basis respectively.

$$
E\left(B_{x,t:c}^{k,C}|\mathcal{F}_c\right) = \sum_{\varsigma\in\{M,F\}} \sum_{i=0}^{2} \left[\frac{\tilde{V}_{x,\varsigma,t:c}^{i,C}}{V_{x,\varsigma,t:c}^{i,C}} \sum_{j\in\mathcal{D}_{i,k},g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k}\, cp_{x,\varsigma,g,t}^{j,C} A_{x,\varsigma,g,t:c}^{j,C} \right],
$$
(3.53)

and

$$E\left(I_{x,t:c}^{k,C}|\mathcal{F}_c\right) = \overline{\Pi}_{x,t}^{k} \sum_{\varsigma\in\{M,F\}} \sum_{i=0}^{2} \left[\frac{\tilde{V}_{x,\varsigma,t:c}^{i,C}}{V_{x,\varsigma,t:c}^{i,C}} \sum_{j\in\mathcal{D}_{i,k},g\in\mathcal{G}} \eta_{\varsigma,g|x,t}^{j|k} {}_c p_{x,\varsigma,g,t}^{j,C} a_{x,\varsigma,g,t:c}^{j,C}\right], \qquad (3.54)$$

for benefits and regular premiums respectively.

Once a life develops dementia, their genotype or whether they entered the insurance pool after having recognised the signs of dementia are irrelevant to their future health status. Additionally, we assume that at a dementia claim, the insurer also learns the life's current functional disability type, making initial functional disability type redundant if no previous claim had been made. The insurer will not observe the life's transition into a demented state until a claim commences, at which point they know how much has already been claimed and what state the life is in. The future value of benefit and premium for lives in the D period can be calculated using the actual mix of business sold:

$$E\left(B_{x,t:c}^{k,D}|\mathcal{F}_c\right) = \sum_{j\in\mathcal{U}_k,g\in\mathcal{G},\varsigma\in\{M,F\}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k} {}_c p_{x,\varsigma,g,t}^{j,D} A_{x,\varsigma,g,t:c}^{j,D}, \qquad (3.55)$$

and

$$E\left(I_{x,t:c}^{k,D}|\mathcal{F}_c\right) = \overline{\Pi}_{x,t}^{k} \sum_{j\in\mathcal{U}_k,g\in\mathcal{G},\varsigma\in\{M,F\}} \tilde{\eta}_{\varsigma,g|x,t}^{j|k} {}_c p_{x,\varsigma,g,t}^{j,D} a_{x,\varsigma,g,t:c}^{j,D}, \qquad (3.56)$$

for benefits and premiums respectively.

To allow for changing business patterns caused by the highest risks buying early, in calendar year c, we assign a relevancy factor based on the age of business sold at time t, $\theta(c-t)$. For its sigmoid shape, we choose an adaptation of the Gompertz function. Reversing the direction of the characteristic 's' shape will allow us to assign the most recent observations a similar, high relevancy; while policies sold long ago are assigned lower relevancy:

$$\theta(c-t) = \exp\left[\beta\exp\left(-\gamma(c-t)\right)\right], \qquad (3.57)$$

where $\beta, \gamma \leq 0$.

We now have everything required to calculate the regular premium adjustment factor $\overline{\psi}_{x,c}^{k}$ as,

$$\overline{\psi}_{x,c}^{k} = \frac{\sum_{y=0}^{c-1} \overline{\psi}_{x,y}^{k} \int_{y}^{y+1} \theta(c-t) \frac{E\left(B_{x,t:\lfloor t\rfloor}^{k,P}|\mathcal{F}_c\right)}{E\left(I_{x,t:\lfloor t\rfloor}^{k,P}|\mathcal{F}_c\right)} dt}{\int_{y}^{y+1} \theta(c-t)\, dt}, \qquad (3.58)$$

and similarly for the single premium factor $\psi_{x,c}^{k}$.

This technique could also be used by regulators to monitor the level of premiums.

It is important that premiums are sufficient such that the company selling LTC will not have issues in solvency.

3.5   Discussion

The Markov property allowed us to easily implement a pricing model which used the development of claims experience to update a pricing basis for adverse selection. However, this property is a simplifying assumption for mathematical convenience required due to limited data being available to fit anything more complex. Without this assumption, the expected present values $A_{x,\varsigma,g,t:c}^{j,\rho}$ and $a_{x,\varsigma,g,t:c}^{j,\rho}$ will be dependent upon the time of the event which the insurer observes. With a simulation based approach like ours, that would make it necessary to store an inhibitively large quantity of data. Additionally, the formulae given in Equations (3.53 to 3.56) would be invalidated and need amending.

Although it is plausible that the Markov property does not reflect reality, our approach gives a first look at a methodology which could be developed if further data were available.

3.6   Summary

We have set out a modelling methodology that will allow us to calculate the benefit outgo and premium income on LTC business, under a set of assumptions about buying behaviour. This came in two stages, first the simulation of health of insured lives to calculate occupancy probabilities and expected present values of benefit outgo and premium income at the start of each calendar year. Secondly, using the results of these simulations, we calculate the premium rates which would be charged if an insurer repriced annually to take into account the information it learns about the mix of existing business from its claims history.

To perform this modelling, we will require additional assumptions over the benefits related to particular states, and of an initial distribution of lives. These are specific to the market to which the model is being applied, so have been purposely excluded from this chapter to retain generality. With such a set of assumptions, the profits or losses which will ultimately be made on business sold at a particular time can be calculated as the difference between the EPVs of benefit outgo and premium income, on the experience basis. The loss which ultimately occurs will represent the cost of adverse selection, and we can observe how this changes in respect of our dynamic premium rates. In the next chapter we do exactly this in the U.K. market.

Chapter 4

Long Term Care Costs in the United Kingdom

In this chapter, we take the LTC models of the previous chapter and run them with appropriate assumptions for economic parameters and buying behaviour. We aim to measure the adverse selection risk in starting up a start-up market for LTC. Furthermore we will perform some sensitivity analyses to consider how sensitive profitability is to our chosen parameters.

4.1   Long-Term Care in the United Kingdom

Although individuals face a risk in their future care costs which might be mitigated by some form of LTC product (Guillén and Comas-Herrera, 2012 presents a methodology for measuring how well this is done), there is currently no deferred-needs cover — where the customer is not currently in need of care — for sale in the U.K. During the 1990s, attempts were made by the U.K. insurance industry to produce and market such a product, but this sold in very few numbers and companies stopped selling it. At the end of 2009, there were about 36,000 LTC policies in force among Association of British Insurers (ABI) members (Commission on Funding of Care and Support, 2010).

In the U.K. there is an expectation among the public, that the state pays for the care of the elderly — Fletcher (2011) quotes an Ipsos-Mori survey that says 54% of the public think care is free at point of use — but in England, Northern Ireland and Wales, personal care is paid for by the individual. However, in these jurisdictions, to protect the poorest, means-tested support is provided by local authorities for those with assets worth under £23,250. The value of one's home can be considered in this financial assessment if a partner no longer lives there, creating the possibility of having to sell one's home to pay for care.

The arrangements in Scotland are quite different. A fixed contribution toward

the cost of care is paid for by the Scottish Government, regardless of an individual's assets. However, the individual is still required to pay 'hotel' costs — the cost of accommodation in a care home — although means-tested support is available for this also. Where the price charged for care is higher than the government's contribution, the liability to pay this is the individual's.

In July 2010 the Commission on Funding of Care and Support, chaired by Sir Andrew Dilnot, began reviewing the system for care support in England and published its findings in 2011. Among the suggestions was a cap on the lifetime care costs of between £25,000 and £50,000 (Commission on Funding of Care and Support, 2011), while hotel costs would not be subject to this and would remain uncapped. The U.K. government has instead proposed to cap care costs at £75,000 in 2017[1], while increasing the point at which means-tested support is made available to assets under £100,000. There has been no mention how inflation will affect the size of the cap but for the purposes of our modelling we assume it behaves as described in Section 3.3.2 (whereby the cap inflates at the same rate as costs and all comparisons between costs and caps are made on the same purchasing power). It is worth noting that other countries have instituted controls over public care provision *e.g.* in France it is only available to those above age 65, and in Spain the first two years of disability are not covered.

Despite the failure of previous attempts at establishing an LTC market in the U.K., Commission on Funding of Care and Support (2011) outlines a role for insurers alongside the state. They investigated whether the problem of pooling of risk could be left solely to the private sector but there was too much uncertainty for affordable products to be designed. They also considered how a fully social scheme could work and found it could be unsustainable and sensitive to political pressure — countries which have adopted such a policy have made cutbacks in response to fiscal pressures resulting in rising unmet needs. The proposition of a cap reduces the 'tail risk' of an insurance policy as this cap transfers part of the risk of excessively large and long claims to the government.

A previous commission, established in 1997, had recommended in 1999 that both personal and nursing care be provided by the state on the basis of need (Royal Commission on Long Term Care, 1999). Moreover, they suggested that private insurance would be too costly. The Westminster government rejected the recommendation of personal care provision but accepted that nursing care should be provided to those who needed it (the Scottish government accepted both recommendations).

Lloyd (2011) suggests various barriers to the success of an LTC market, from the perspective of both the supply and demand sides. Among the supply barriers is adverse selection costs (recall from Chapter 1 that there was evidence for adverse

---

[1]Hansard HC Deb 11 February 2013, vol 558, cols 592–607.

selection in LTC products found by Finkelstein and McGarry, 2006; Courbage and Roudaut, 2008; Oster et al., 2010). In such a small volume of business as the U.K.'s, the inability to spread the cost of high claims could pose a significant risk to the insurer. It should be noted that Finkelstein and McGarry (2006)'s adverse selection did not translate into high claims, possibly due to an incorrect self-assessment of the survey participants' health. Other factors he discussed were:

- Limited profitability and market size — weighing the potential costs of entering the market against the potential profits, it might not be seen worth re-entering the market.

- Longevity and morbidity risk — there is significant uncertainty over the future of life expectancies and how trends in disability might be shaped in the future.

- Uncertainty over future claims patterns — changes in availability of informal care or social care services in response.

- Reputation risk — if benefits do not meet the needs of customers, there could be an effect on sales for other products so an insurer may need to be generous in terms of claims underwriting.

- Financial advisors' resistance — the product history may be tainted in the eyes of advisors making it difficult to sell through these channels.

- Claims assessments — providers of care may be incentivised to overstate needs knowing that the cost was being covered by an insurer. Training underwriters to perform full claims underwriting would be a significant up-front cost. However, evidence from the health insurance industry in the U.S. suggests the uninsured would get overcharged since insurers have greater negotiating power (Melnick and Fonkych, 2008).

- Solvency II — capital requirements with such uncertainty could be inhibitive to entering the market.

From the demand side he lists reasons why an individual might not buy the product. The key reasons are:

- Cost — many households would have great difficulty in affording the large single premiums on a LTC contract, especially in the context of competing expenditure requirements.

- Alternative strategies — some households keep their savings or plan on using anticipated inheritance from older relatives or the value of their home in order to pay for the old age requirements.

- Ignorance over the need for care arrangements — as mentioned above, many do not realise the care system is not fully state supported. Additionally, there is some ignorance as to the in-home care requirements some people will experience in their old age.

- The complexity of products and a distrust of financial services.

Furthermore, Pauly (1990) makes the case for the rational non-purchase of LTC based on utility theory and a lack of awareness to the probability that long-term care services are required. He considers where utility can and cannot be gained from a bequest and finds that maximising expected utility might mean not purchasing the insurance, although this is in the specific context of the U.S. Medicare programme. Additionally, he argues that the purchase of LTC disincentivises the provision of informal care from children which would allow them to stay in their own home rather than moving to a care home.

Zhou-Richter et al. (2010) used empirical data to investigate the role of adult children and their knowledge of care costs in the demand for LTC. They surveyed 1045 adult children in Germany, asking about ownership of LTC by themselves and their parents and after giving various statistics of the costs and probabilities of requiring long-term care services, whether they were more or less likely to purchase cover. Their results suggested that the increased knowledge among adult children about the likely needs of their elderly parents, leads to increased willingness to buy insurance. To reinforce this idea, they looked at whether the decision to buy insurance of those who already had it had been influenced by knowledge of the risk and their results supported such a link.

4.2 Parameterising the UK LTC Market

In this section we estimate the market specific parameters necessary for the modelling described in Chapter 3. First we will parameterise the cost of care provision in each health state and secondly, the initial distribution of lives at 1st January, 2013.

4.2.1 Care Costs

The costs to pay for care varies greatly by region. Region specific costs for care homes, nursing homes and hourly rate for in-home carers are given in Tables 4.1 and 4.2. The in-home care cost differs for weekend and weekday (as well as by day or night but for simplicity we assume daytime) and we find the average daytime rate as Weekday rate $\times 5/7$ + Weekend rate $\times 2/7$. To find the UK average for each, we calculate the weighted average using the 2010 population size as weights from Office

Table 4.1: Average annual costs (£) for a residential care home with and without nursing by UK region in 2011/12. Source: Laing & Buisson, Care of Elderly People Report 2012/13 via `http://www.payingforcare.org`

Region	Care Home Fees	
	With Nursing	Without Nursing
East Midlands	32,136	26,312
East of England	41,600	29,328
London	42,692	31,096
North East	31,044	24,492
North West	34,476	24,336
Northern Ireland	29,640	24,232
Scotland	35,620	28,860
South East	45,188	30,888
South West	39,728	28,652
Wales	33,800	25,532
West Midlands	36,816	25,740
Yorkshire & Humber	32,448	24,076

for National Statistics (2012a) given in Table 4.3. The costs need to be inflated to our modelling start date, 1st January 2013. Care home costs are given for the period 2011/2012 so we assume they are as at 1st January, 2012 while in-home care costs are for the period 2009/2010 so we assume they are as at 1st January, 2010. Office for National Statistics (2013) gives sector specific weekly earnings up to November, 2012, from which we calculate salary inflation. In-home care costs are inflated using the experienced inflation rates for health and social workers to 1st January, 2012. An annualised rate for January, 2012 to November, 2012 is assumed to cover the full year from 1st January, 2012 to 1st January, 2013 and is applied to inflate all care costs to this date. Thus the average annual costs for staying in a care home with and without nursing provision are £27,900 and £38,100 respectively. The average hourly rate for in-home care is £14.00.

Once the government's cap has been reached, the individual has no liability for care costs, but still must pay for staying in a care home, for food and for the utilities — the so-called hotel costs. This cost is fixed regardless of whether there is any nursing care provided. Hancock et al. (2007) breaks up the hotel costs from the full cost of staying in a residential care home without nursing care in 2002 as £7,900 *per annum* and £17,000 *per annum* respectively. They based their estimate of hotel cost on the Guarantee Credit component of the U.K. Pension Credit benefit. We assume the proportions remain the same today, *i.e.* hotel costs are 46% of residential care home without nursing care, or £12,965.

We consider an institutionalised life without any ADLs or with only an IADL as

Table 4.2: Average hourly daytime in-home care cost (£) by UK region in 2009/10. Source: Laing & Buisson, Domiciliary Care UK Market Report 2011 via `http://www.payingforcare.org`

Region	Weekday	Weekend	Average Day
East Midlands	12.79	14.66	13.32
East of England	14.02	15.12	14.33
London	13.94	16.41	14.65
North East	13.94	14.88	14.21
North West	12.13	12.48	12.23
Northern Ireland	10.44	10.44	10.44
Scotland	11.42	12.7	11.79
South East	13.61	15.16	14.05
South West	13.98	14.9	14.24
Wales	11.91	13.65	12.41
West Midlands	12.34	13.74	12.74
Yorkshire & Humber	13.23	13.62	13.34

Table 4.3: Population sizes of UK regions in 2010. Source Office for National Statistics (2012a)

Region	Population
East Midlands	4,481,431
East of England	5,831,845
London	7,825,177
North East	2,606,625
North West	6,935,736
Northern Ireland	1,799,392
Scotland	5,222,100
South East	8,523,074
South West	5,273,726
Wales	3,006,430
West Midlands	5,455,179
Yorkshire & Humber	5,301,252

requiring only residential care. Lives who are in an institution and suffering more than 1-ADL will need the most care so we assign these lives to a residential care home with nursing care. Institutionalised lives with 1-ADL might need a bit less attention but more than lives with no functional disability, so we arbitrarily deem this halfway between the cost of a residential care home with and without nursing care.

The condition for claiming based on functional ability in the LTC market is commonly to suffer from at least 2 ADLs. The costs faced by these lives depends on the number of hours of in-home care required. Jones (2006) surveyed 28 care providers (most of which were in a single local authority area) to find the number of hours of formal in-home care provided. They found an average of 6.5 hours per week (from a total of 6442 hours weekly). This only measures the formal care; informal care which might be provided by the life's spouse, child or a friend is excluded.

Forder and Fernández (2009) found that the local authority's care provision depended greatly on whether the life lived alone or had access to informal care. They assumed a need based on what is received by lives eligible under the Department of Health's Fair Access to Care framework: 1 ADL (6.8 hours per week); 2 ADLs (7.3 hours per week); 3 ADLs (8.8 hours per week); 4 ADLs (15.6 hours per week); and 5 ADLs (18.7 hours per week). If the life is insured, when they meet the claim requirement, they would be contractually entitled to benefits regardless of their living arrangements if they chose to get a carer. In this situation, with lives who (wrongly) think they are being a burden on their family, the number of hours of care could be substantially more than what the local authority faces. There could however be a stigma felt by claimants not wanting to be seen by their peers as needing a professional carer which might curb this increase slightly.

Nuttall et al. (1994)'s model assumed a number of hours per week which differed by level of disability and classified by need: low need (5 hours per week); moderate need (15 hours per week); regular need (30 hours per week); and continuous (45 hours per week). We leave the number of hours a week as a scenario specific parameter and denote it by H.

For our contract, when a life has been diagnosed with dementia, this will reduce the functional disability requirement to trigger a functional disability claim to 1-ADL. Lives with dementia and > 1ADL are given $2H$ hours of care a week.

Without the existence of the government cap, payments faced by the individual which don't meet the claim requirements are irrelevant to the insurer. In our market, if the individual pays for care which does not meet the insurance company's claim underwriting but is included within the government's cap, this will reduce the insurer's expected future liability. To simplify the situation, we assume the government and insurer have the same criteria regarding claims and set the care costs in states other than those detailed above to be zero.

Table 4.4: Annual rates for care and hotel costs for individuals in our insurance market.

State Index	State Name	Care Cost (£)	Hotel Cost (£)
0	Healthy	0	0
1	IADL	0	0
2	1-ADL	0	0
3	> 1ADL	$5114H$	0
4	Initial Signs	0	0
5	IADL(IS)	0	0
6	1-ADL(IS)	0	0
7	> 1ADL(IS)	$5114H$	0
8	Dementia	0	0
9	IADL(D)	0	0
10	1-ADL(D)	$5114H$	0
11	> 1ADL(D)	$10227H$	0
12	Institution	14935	12965
13	IADL(I)	14935	12965
14	1-ADL(I)	20035	12965
15	> 1ADL(I)	25135	12965
16	Dead	0	0

A summary of the care costs dependent on the current state in the model depicted in Figure 3.7, are given in Table 4.4.

4.2.2 Initial Distribution of Lives

Unlike our model of an established CI market in Chapter 2 where we modelled from age 30 and assumed all lives started healthy, in our start-up market, we model the progression of the market from introduction of the product. In such a case there will be lives of different ages buying the product and influencing the adverse selection costs. Some lives at each age have already developed some form of cognitive or functional disability so to model forward from the current time, it is necessary to calculate the distribution of lives in each state in our model for each sex, genotype and age group as at 1st January, 2013. To do this, we will first estimate a mix of lives at age 60. From this mix at age 60, we will calculate the occupancy probabilities for each state for the elderly population which we will then use to form our insurance market, split into 5-year age groups in terms of age last birthday: 60–64, 65–69,70–74, 75–79, 80–84 and 85–89.

Denote the probability that a life has genotype $g \in \mathcal{G}$ at age 60, by $P_{\mathcal{G}}(g)$. Further denote the probability that a life of sex ς, has functional disability type $i \in \{0, 1, 2, 3\}$ at age 60 by $P_{\text{ADL},\varsigma}(i)$.

Table 4.5: Distribution of APOE genotypes. Source: Farrer et al. (1997).

Genotype	Probability, $P_{\mathcal{G}}(g)$
$\varepsilon 2 \varepsilon 2$	0.135
$\varepsilon 2 \varepsilon 4$	0.026
$\varepsilon 3 \varepsilon 3$	0.609
$\varepsilon 3 \varepsilon 4$	0.213
$\varepsilon 4 \varepsilon 4$	0.018

Since we are concerned with dementias of old-age, we assume that no lives have any signs of decrease of cognitive function at age 60 and the distribution of genotypes is that used by Macdonald and Pritchard (2000), and derived by Farrer et al. (1997) (shown in Table 4.5). Farrer et al. (1997) performed a meta-analysis of 40 studies of APOE, and provided the genotype frequencies among both lives with AD and the controls split by ethnicity (the chosen results being Caucasian controls).

The remaining part of the mix of lives at age 60 exactly, is the prevalence of functional ability. We find the prevalences of functional ability at age 60 for males and females, which when projected forward following the transition intensities (for simplicity, we exclude mortality improvements) in our model of health (see Figure 3.7), best fits the prevalences observed in the first phase of the CFAS study, as reported by Akodu (2007) (see Tables 4.6 and 4.7). They list these at age groups in terms of age last birthday, 65–69,70–74,75–79,80–84,85–89,90+, which we represent by the set of midpoints (and 92.5 for 90+), denoted $\mathcal{X}'$, where $\mathcal{X}' = \{67.5, 72.5, 77.5, 82.5, 87.5, 92.5\}$. By varying the proportions originally in each state at age 60, we use a weighted least squares method, with weights equal to the number of lives in each age group, as estimated by Office for National Statistics (2011) for 2010 (shown in Table 4.8), to minimise the difference between the observed and our calculated prevalences, thereby extrapolating Tables 4.6 and 4.7 to age 60. The function to minimise is,

$$\sum_{x\in\mathcal{X}'}\left[\sum_{j\in\{0,1,2,3\}}\left(Obs_x^j - \frac{\sum_{g\in\mathcal{G},i\in\{0,1,2,3\},k\in\{0,4,8,12\}}{}_{x-60}p_{60,\varsigma,g}^{i,j+k}P_{\mathcal{G}}(g)\,P_{\mathrm{ADL},\varsigma}(i)}{\sum_{g\in\mathcal{G},i\in\{0,1,2,3\},k\in\mathcal{S}\setminus\{16\}}{}_{x-60}p_{60,\varsigma,g}^{i,k}P_{\mathcal{G}}(g)\,P_{\mathrm{ADL},\varsigma}(i)}\right)^2\right]P_{\mathcal{X},\varsigma}(x),$$
(4.1)

where Obs_x^j is the observed prevalence of functional ability j, at age x, subject to the constraints that $\sum_{i\in\{0,1,2,3\}}P_{\mathrm{ADL},\varsigma}(i)=1$, and $P_{\mathrm{ADL},\varsigma}(i)\in[0,1]$. The occupancy probabilities, ${}_{x-60}p_{60,\varsigma,g}^{i,j+k}$, are found by solving the Kolmogorov forward equations,

$$\frac{d}{dt}{}_tp_{x,\varsigma,g}^{i,j} = \sum_{k\neq j}\left({}_tp_{x,\varsigma,g}^{i,k}\,\mu_{x,t,\varsigma,g}^{k,j} - {}_tp_{x,\varsigma,g}^{i,j}\,\mu_{x,t,\varsigma,g}^{j,k}\right).$$
(4.2)

Table 4.6: Distribution of functional ability in males by age group in terms of age last birthday. Source: Akodu (2007).

Functional	Age group					
Ability	65–69	70–74	75–79	80–84	85–89	90+
None	0.8406	0.8112	0.6791	0.5382	0.3545	0.1757
IADL	0.0534	0.0602	0.0756	0.1191	0.1231	0.1081
1-ADL	0.0737	0.0860	0.1556	0.2075	0.3134	0.2568
> 1ADL	0.0323	0.0426	0.0898	0.1352	0.2090	0.4595

Table 4.7: Distribution of functional ability in females by age group in terms of age last birthday. Source: Akodu (2007).

Functional	Age group					
Ability	65-69	70-74	75-79	80-84	85-89	90+
No ADLs	0.7759	0.6854	0.5486	0.3483	0.1813	0.0847
IADL	0.1013	0.1194	0.1382	0.1749	0.1827	0.1229
1-ADL	0.0851	0.1515	0.2272	0.302	0.3626	0.2542
> 1ADL	0.0376	0.0436	0.0860	0.1749	0.2734	0.5381

This was done using a 4th order Runge Kutta algorithm with a step size of 2^{-12} and boundary conditions $_0p_{x,\varsigma,g}^{i,j} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. The result of the minimising Equation (4.1) is shown in Table 4.9, with no lives having any ADLs at age 60.

Using the assumptions and calculated distributions of lives at age 60, we can find the distribution of lives who are alive at age x, of sex ς, have genotype g and are in

Table 4.8: U.K. population by sex and age group in terms of age last birthday in 1,000s. Source: Office for National Statistics (2011).

Age	Male	Female
60–64	1840.08	1923.52
65–69	1412.11	1519.56
70–74	1160.31	1307.44
75–79	893.91	1107.84
80–84	607.09	885.55
85–89	326.08	608.47
90+	131.89	331.54

Table 4.9: Proportion of lives of each sex in each functional disability type, at age 60 exactly.

	No ADLs	IADL	1-ADL	> 1 ADL
Male	1	0	0	0
Female	1	0	0	0

state j at 1st January 2013, $p^j_{x,\varsigma,g}$, by calculating,

$$p^j_{x,\varsigma,g} = \frac{\sum\limits_{i=0}^{3} P_{\text{ADL},\varsigma}(i)\; P_{\mathcal{G}}(g)\; {}_tp^{i,j}_{60,\varsigma,g}\; P_{\text{Sex},x}(\varsigma)}{\sum\limits_{\varsigma \in \{M,F\}, g \in \mathcal{G}, s \in \mathcal{S}} \sum\limits_{i=0}^{3} P_{\text{ADL},\varsigma}(i)\; P_{\mathcal{G}}(g)\; {}_tp^{i,s}_{60,\varsigma,g}\; P_{\text{Sex},x}(\varsigma)}, \tag{4.3}$$

where $P_{\text{Sex},x}(\varsigma)$ is the probability a life aged x is of sex ς based on Table 4.8.

This will not necessarily produce an accurate depiction of the mix of lives of the U.K. population, however the purpose of such a distribution is to provide an approximate baseline from which we can illustrate how adverse selection may impact costs. The aim of our model is not to accurately estimate or project future demand for services and our results should not be used in this way.

4.3 Distribution of Insured Lives

To provide context for adverse selection costs, we consider the joint distribution of insurance purchase time, age and sex, conditional on insurance being purchased. Let T^{INS}, X and Σ be the random variables representing insurance purchase time, age group and sex respectively. We denote the joint density function for insurance purchase time, sex and age group, conditional on insurance being purchased by $f_{T^{INS},\Sigma,X|T^{INS}<\infty}(t,\varsigma,x)$, and calculate it as,

$$f_{T^{INS},\Sigma,X|T^{INS}<\infty}(t,\varsigma,x) = \frac{\sum\limits_{g \in \mathcal{G}, i,j \in \mathcal{S}, \vartheta \in \mathbb{B}} p^i_{x,\varsigma,g}\; {}_tp^{00i,0\vartheta j}_{x,\varsigma,g}\; \mu^{0\vartheta j,\vartheta j}_{x,t,\varsigma,g}}{\int_0^\infty \sum\limits_{y \in \mathcal{X}, \sigma \in \{M,F\}, g \in \mathcal{G}, i,j \in \mathcal{S}, \vartheta \in \mathbb{B}} p^i_{y,\sigma,g}\; {}_sp^{00i,0\vartheta j}_{y,\sigma,g}\; \mu^{0\vartheta j,\vartheta j}_{y,s,\sigma,g}\; ds}, \tag{4.4}$$

with the integration calculated numerically using Simpson's rule and a step size of 2^{-11}. The resulting density functions are shown in Figures 4.1 (males) and 4.2, assuming adverse selection is due to lives with the initial signs of dementia or 1-ADL buying at a rate of 0.25 *per annum*.

We can see that the majority of business is sold during the first 10 to 15 years; very few lives buy when $x + t$ is greater than 90 years old. There is an interesting

feature among both males and females, for $x = 62.5$, whereby there is a local minimum followed by a local maximum for the smaller markets at $t = 5$ and $t = 15$, respectively for males and $t = 2$ and $t = 10$, respectively for females. These turning points are due to the changing pattern of health with age: initially the adverse selectors in this age group are functionally disabled; as dementia onset increases, the adverse selection comes from lives with the initial signs of dementia.

Since we are concerned with a start-up market, we choose to model for 20 years. As very few lives purchase insurance after this time (we have not modelled entry to the market from younger lives), the implication of this is limited. Moreover, this is longer than the previous attempt at selling the product in the U.K. lasted. If the product is successful, a large body of claims history might allow a more robust analysis, perhaps performed by the CMI, to estimate an insured lives morbidity table which insurers could adjust to suit their own market segment.

4.4 Analysis of Premium Rates

In this section we calculate and analyse the premiums that would be charged to lives if the insurer knew all relevant information (applicant's exact health status, sex and genotype) and was able to underwrite fully *i.e.* an underwriting class for each sex, state and genotype combination. This will aid the understanding of the potential sources of adverse selection.

To do so we run the previous chapter's models with assumptions as follows:

- The insurance contract indemnifies the life up to a maximum lifetime coverage of £200,000, with no annual claim limit.

- Care costs, other than hotel costs, are capped at £75,000, adjusted for 4 years of inflation at a force of inflation of $\nu = 0.04$ *per annum*.

- The number of hours of in-home care per day for lives with $> 1\text{ADL}$ is $H = 1$.

- There is no deferred period.

- The force of interest is $\delta = 0.04$ *per annum*.

- The market size *i.e.* the rate at which lives are assumed to buy insurance, is 0.001 *per annum*.

The resulting regular premium rates *per annum*, dependent on purchase time, are shown in Figures 4.3 to 4.8 (note that scales are not consistent from figure to figure). Figure 4.3 shows the progression of premium rates for lives who were aged 62.5 at 1st January, 2013 and purchase their insurance at time t; Figure 4.4 shows the same for
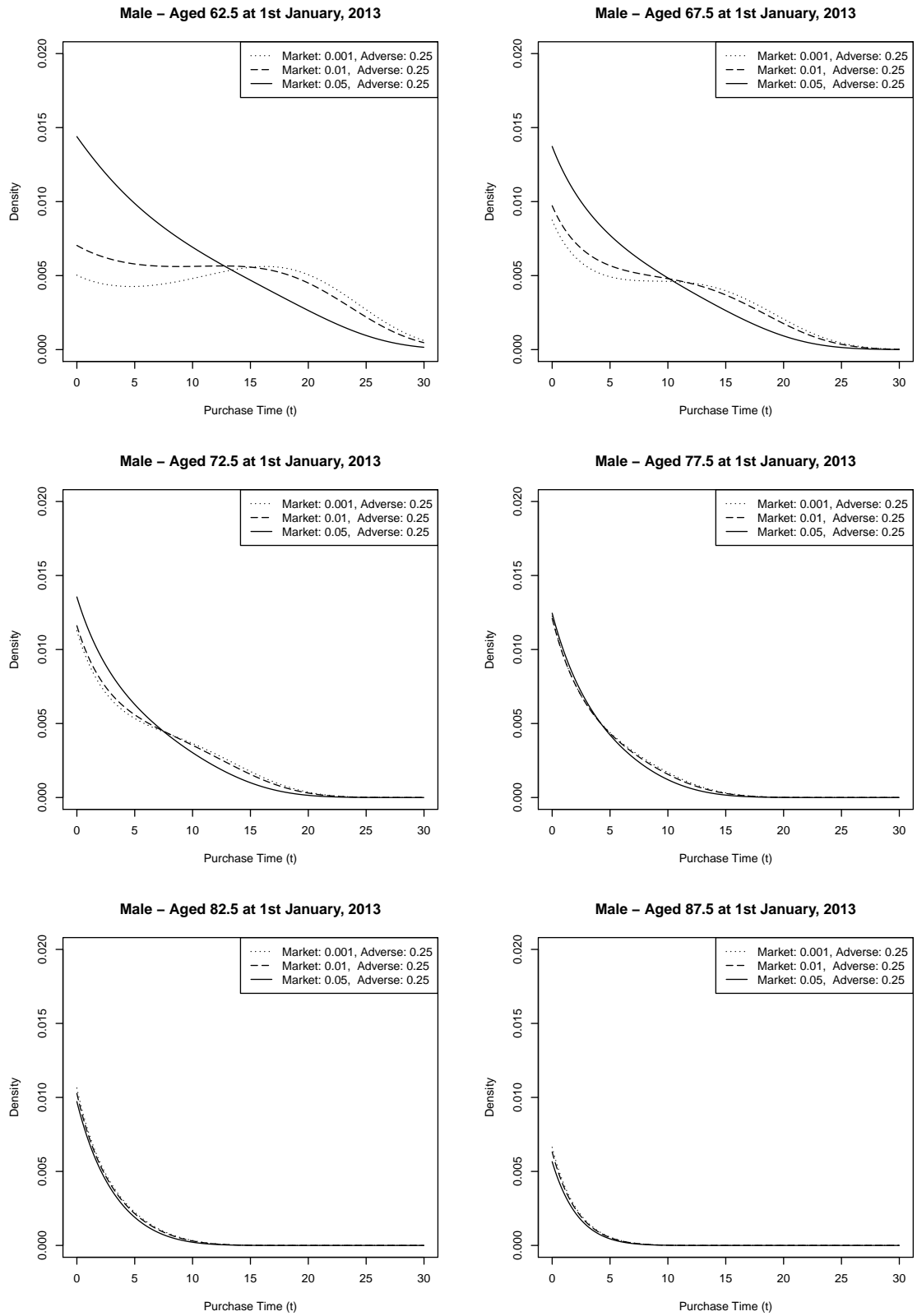
Figure 4.1: Joint density function for insurance purchase time, sex and age group, conditional on insurance being purchased for males.
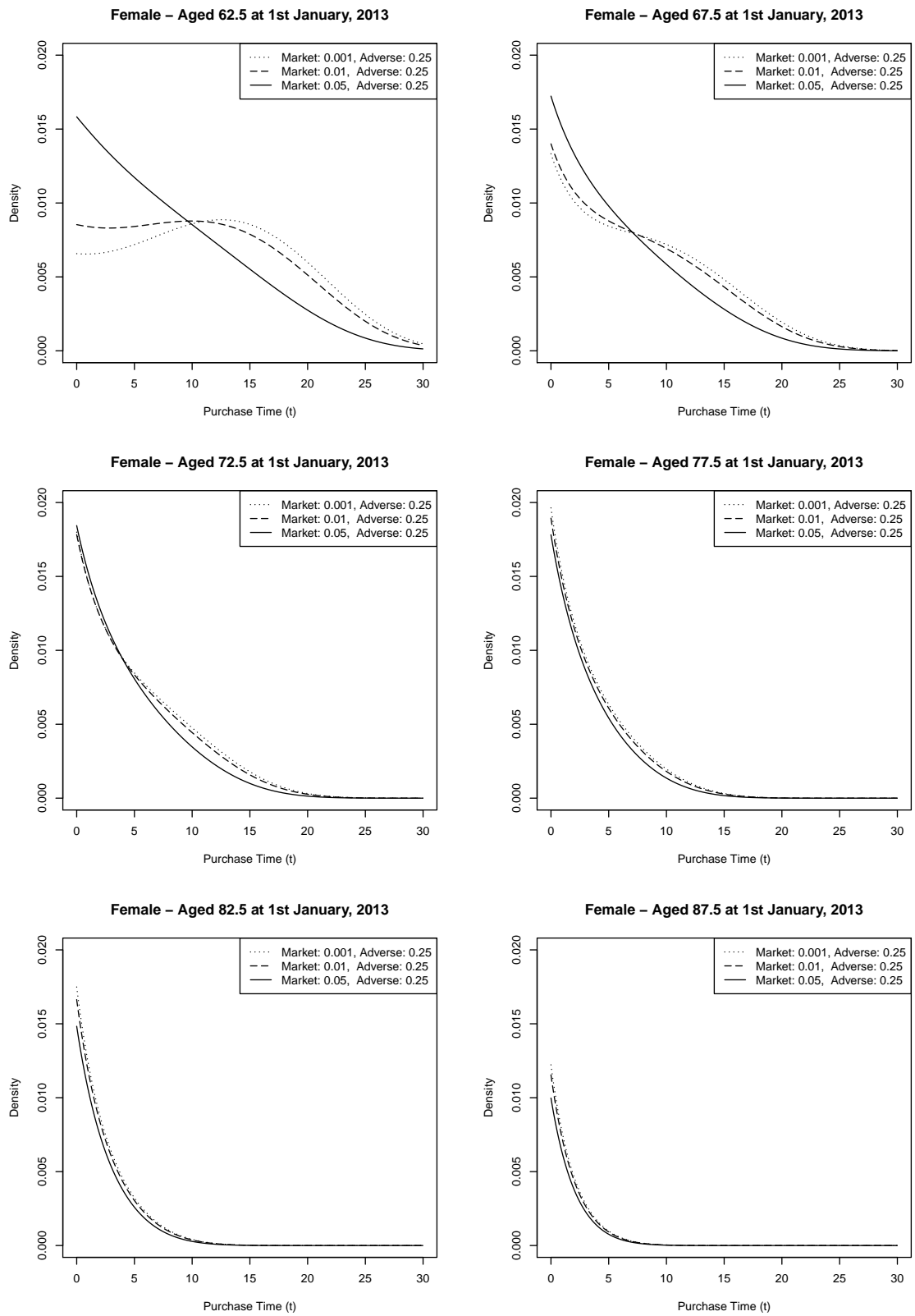
Figure 4.2: Joint density function for insurance purchase time, sex and age group, conditional on insurance being purchased for females.

lives who were aged 67.5 at 1st January, 2013 and so on. Overlayed onto each of these are the premium rates that would be charged if all lives buying insurance were written into one underwriting class for the age group, without any adverse selection in the market (what we termed *base premium* in Section 3.4 and calculated using Equation (3.42)). The equivalent plots for single premium rates are shown in Appendix D

We can see the biggest difference in premium rates is derived from the initial signs of dementia. Also, if pricing men and women separately, premiums for women are larger than those for men, reflecting the higher rate of dementia found in women.

Prior to showing the initial signs of dementia, the premium rates do not differ hugely by genotype. The effect of moving to an initial signs type state increases the premium rates more for females than males. Since this is closer to a dementia claim, the effect of genotype becomes more important and differences between premium rates for genotypes are apparent.

At around age 80, the fully underwritten premium rates generally start to decrease and the differences between them lessen. This is due to the increased mortality related to old-age causing the probability that a life claims to decrease. Additionally, where a life does claim, high mortality acts to cause the length, and hence the value, of claims to be lower. An exception to this observation is made for males in the IADL state. In this case, the proportion of transition rates into $> 1\text{ADL}$ is increasing and hence the probability of a claim is also increasing.

Also at around age 80, the premium rates for genotype $\varepsilon 4 \varepsilon 4$ drop below those for $\varepsilon 3 \varepsilon 3$. At these ages the relative risk for AD, relative to $\varepsilon 3 \varepsilon 3$, is only slightly greater than 1 (for males this is fixed as a minimum), while the integral in Equation (3.28) is greater for $\varepsilon 4 \varepsilon 4$ and hence the transition rates between Initial Signs and Dementia states are greater for $\varepsilon 3 \varepsilon 3$.

After age 85, the premium rates charged start to decrease. When this occurs, the rational behaviour for lives who are not in claim would be to continuously lapse and re-enter for a cheaper premium. To avoid this problem, contract designs would need to have some form of advancement of premium payment *e.g.* by limiting the period for which premium is payable. The adverse selection cost for such a design would be somewhere between that of a single premium contract and our regular premium so we continue without any action to correct or account for this lapse risk.

In terms of perceived cheapness of the premiums that would be charged without full information (the thick lines in the figures), at younger ages when the majority of lives have no signs of dementia, premiums are closer to the Healthy rates, so in 1-ADL it appears cheap. For lives with the initial signs of dementia, base premiums represent as much as half of the premium which they would pay if fully underwritten. As age increases, the proportion of lives with the initial signs of dementia increases and premiums become expensive even for lives with 1-ADL. The high unisex premium
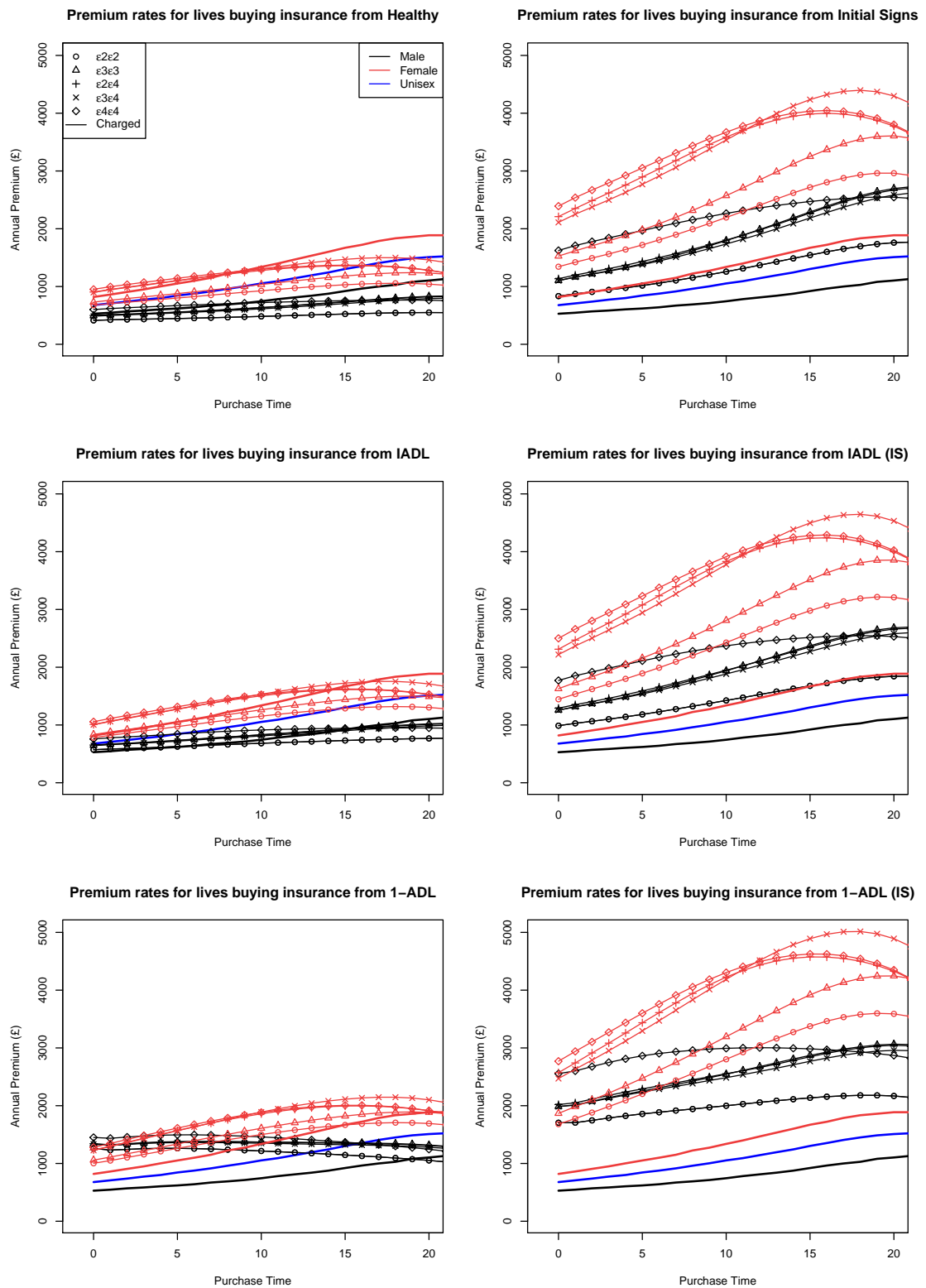
Figure 4.3: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 62.5 at 1st January, 2013. The same legend is used throughout the plots.
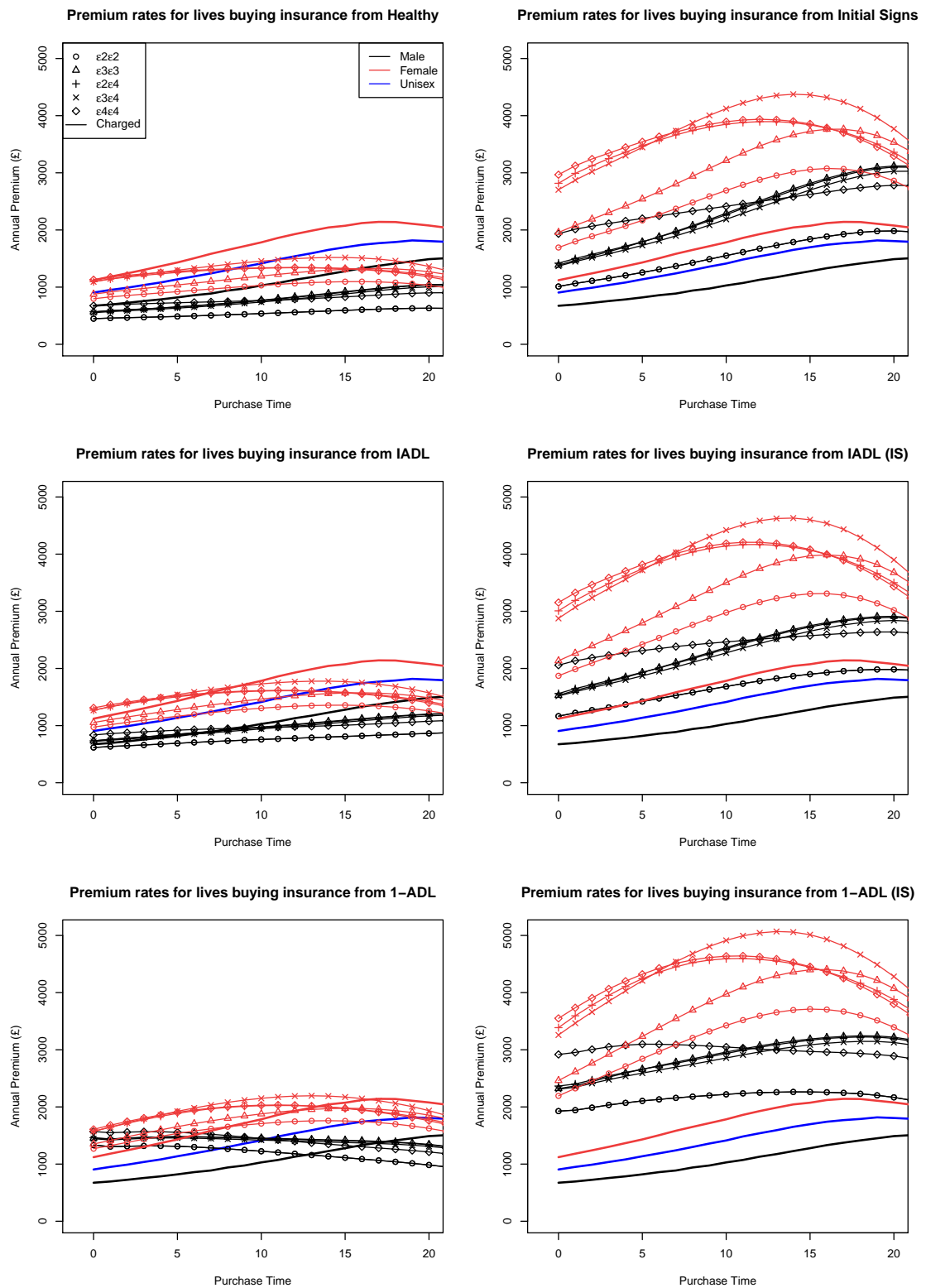
Figure 4.4: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 67.5 at 1st January, 2013. The same legend is used throughout the plots.
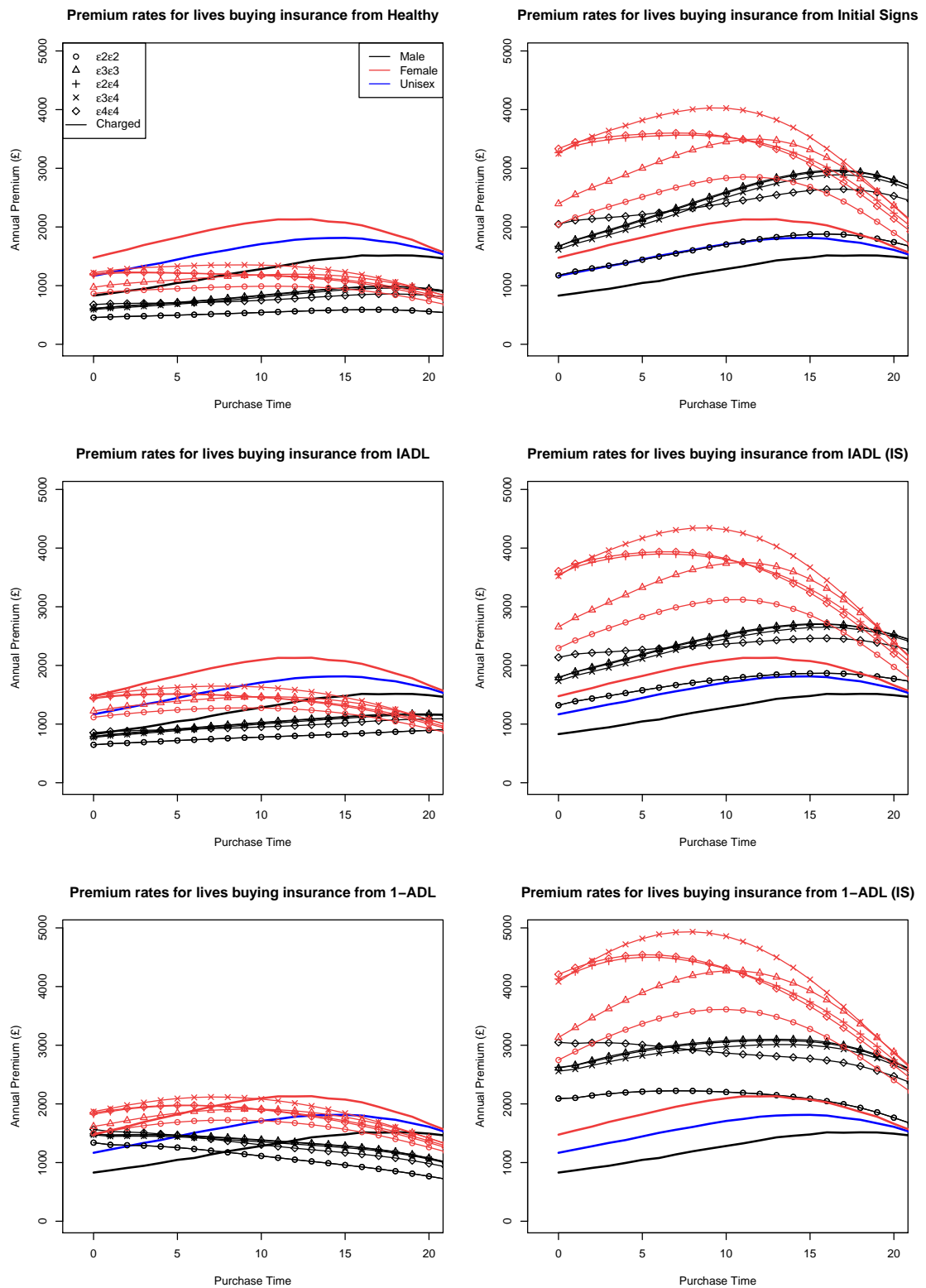
Figure 4.5: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 72.5 at 1st January, 2013. The same legend is used throughout the plots.
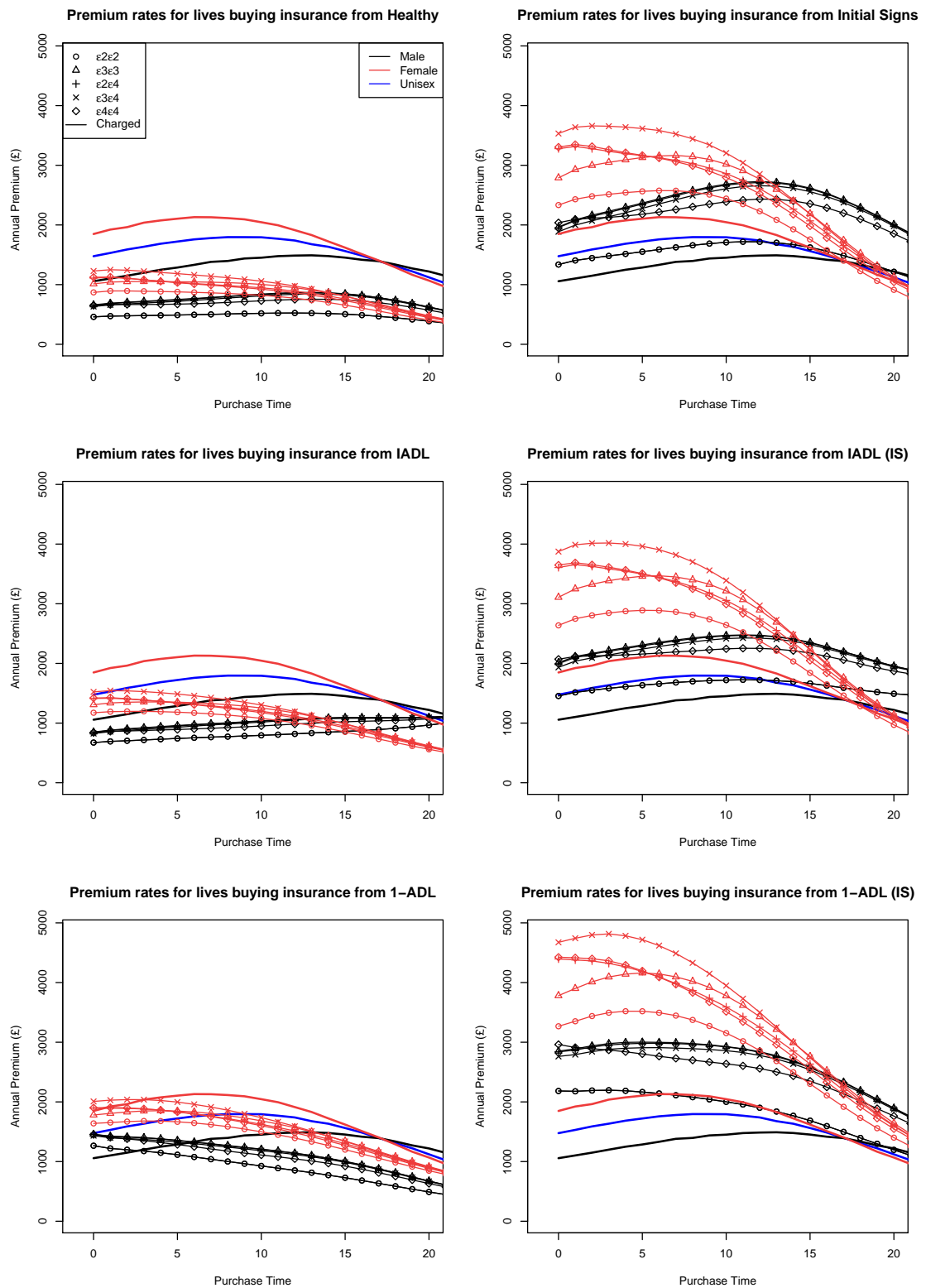
Figure 4.6: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 77.5 at 1st January, 2013. The same legend is used throughout the plots.
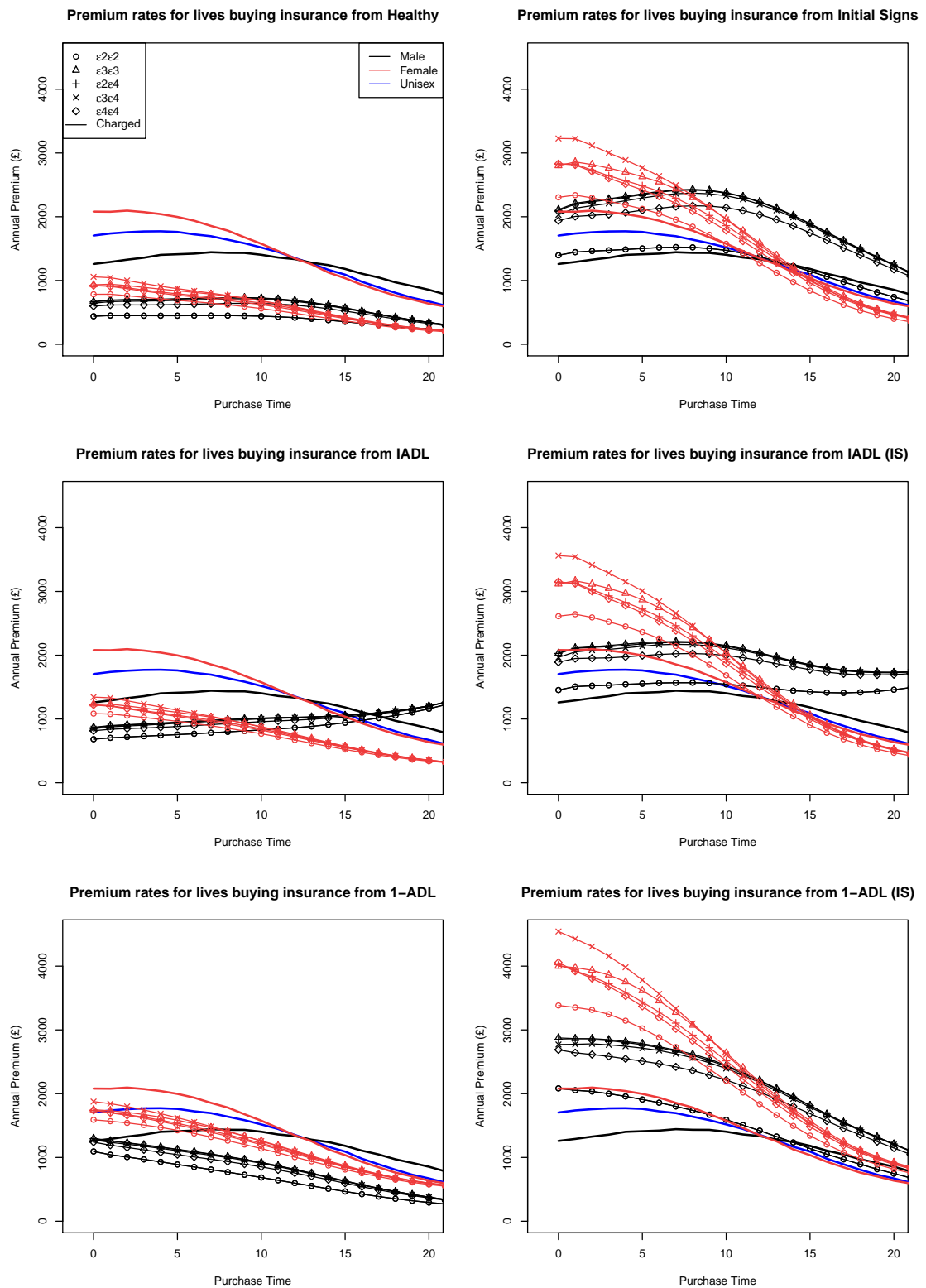
Figure 4.7: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 82.5 at 1st January, 2013. The same legend is used throughout the plots.
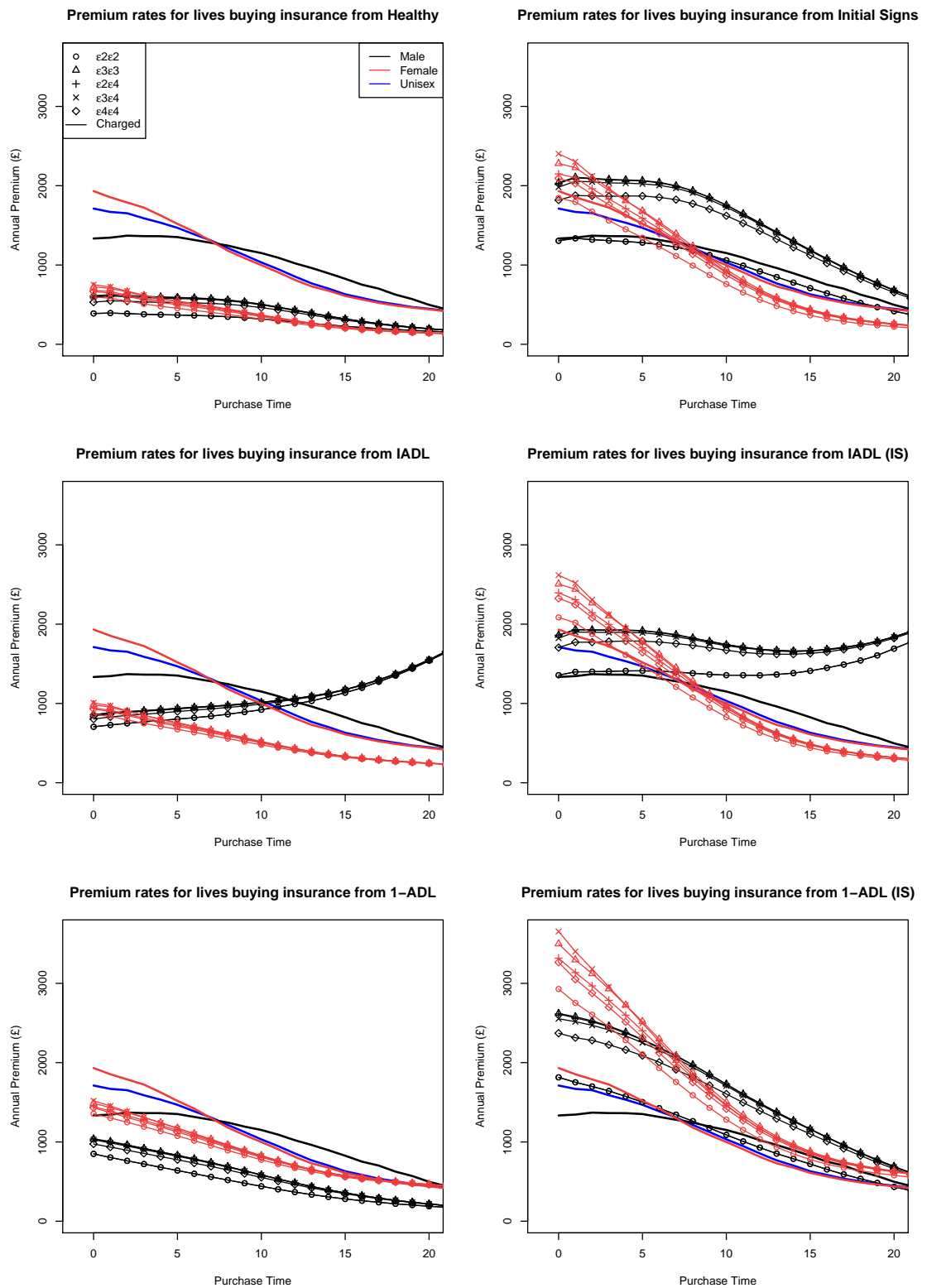
Figure 4.8: Regular premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 87.5 at 1st January, 2013. The same legend is used throughout the plots.

caused by female lives' dementia claims makes the premium seem expensive for males with APOE genotypes $\varepsilon 2\varepsilon 2$, $\varepsilon 2\varepsilon 4$ and $\varepsilon 3\varepsilon 4$. Where people see the premium as cheap, we assume they buy the insurance at a higher rate as they are receiving better value for money and this we see as adverse selection.

Attention should be drawn to the people who these products would be marketed to: old people who own their own home, but nonetheless are living on a pension. Premiums between between £1,500 and £2,000 *per annum* are likely to be seen as a substantial portion of a tight budget.

4.5   Calculating Adverse Selection

In this section we consider the impact of lives making the rational decision to either buy LTC insurance at an increased rate when faced with a cheap premium with respect to what they know about their own risk or at a lower rate when premium is expensive. This will be calculated in the context of a new market for the product being created at the modelling date, 1st January, 2013. We will show how the cost of adverse selection, as defined in Section 2.5.1, will develop over time when the insurance company is able to respond to the cashflows it observes from in-force business by adjusting premium rates charged to new business. Each age group is modelled independently of any other so the repricing adjustments relate only to the claims history of the particular cohort.

Based on the analysis in Section 4.4, we consider the possibility of higher than expected purchase rates from lives who find they have a high risk variant of the APOE gene; lives who have observed the initial signs of dementia; and lives in a high risk ADL state beyond the detection of underwriting capabilities in the scenario. We also allow lives with a low risk variant of the APOE gene or in the lowest risk state in the class to buy at a lower rate.

4.5.1   Adverse Selection Sources in Isolation

To demonstrate the relative impact of each source of adverse selection, we model with buying behaviour changing due to one factor at a time. We use the same set of assumptions as above and where the buying rate is high, this is at a rate of 0.25 *per annum*.

For the purposes of this exercise, we consider high risk variants of APOE to be $\varepsilon 3\varepsilon 4$ and $\varepsilon 4\varepsilon 4$, while low risk variants are $\varepsilon 2\varepsilon 2$ and $\varepsilon 3\varepsilon 3$ with testing occurring at a rate of 0.08916 *per annum*, the High rate from Section 2.5. This is used to exaggerate the costs in order to illustrate how genetic testing interacts with other scenarios. As discussed above, testing for APOE variations in relation to AD is unlikely to be done at the request of physicians, but may be done if an individual uses a personal testing

service. Figure 4.9 shows the repricing adjustment factors to the base premiums over time, using Equation (3.58) — these are calculated based on the information the insurer gains through its claims history. To provide this with context, Figure 4.10 shows the adjustments that would be made to the base premiums, if the insurance purchase pattern was exactly the same, and the insurer knew everything about the customer which was relevant to pricing when the policy was sold — this is the factor by which the base premium should be multiplied to pay for all benefit outgo.

What we can observe from these plots is that lives with the initial signs of dementia buying at a higher rate should have the biggest effect on premium rates, although this is overtaken at the oldest ages by lives with 1-ADL. However, the rate at which losses from this source of adverse selection are recognised is slower than that of lives in the Healthy state not buying and lives with 1-ADL buying at a high rate. This means the premiums charged to new business don't increase enough to cover the benefit costs of the business, leaving a larger cost of adverse selection.

There is a general downward trend in the adjustments that should be made as the relative difference between the underlying premium rates decreases. This is caused by the higher proportions of lives with dementia or functional disability among the population aligning the base premiums more closely to the adverse selectors.

We can also see that on its own, genetic testing at this test rate should increase premiums a moderate amount, however the threat posed by this is small relative to the other potential sources of adverse selection. Overlap between the factors may cause smaller influence from genetic tests than these would suggest. A limitation of the methodology used — lives with particular genotype buy at a higher rate for the remaining period after receiving test results — is that at later ages, the relative risk model used gives the "high risk" APOEε4 variants, $\varepsilon 2\varepsilon 4$ and $\varepsilon 3\varepsilon 4$ the same relative risk as $\varepsilon 3\varepsilon 3$ in males.

4.5.2 Multiple Sources of Adverse Selection

In assessing a more complete picture of adverse selection, with more than one source at a time, we will use the following assumptions:

- The size of the market will be represented by purchase rates of 0.001, 0.01 and 0.05 *per annum*.

- Lives who buy at an increased rate will do so at either 0.25 or 0.1 *per annum* and in the case of the smallest market, at 0.01 *per annum* also.

- Underwriters will either group all lives into one class or group lives with 1-ADL in their own class.
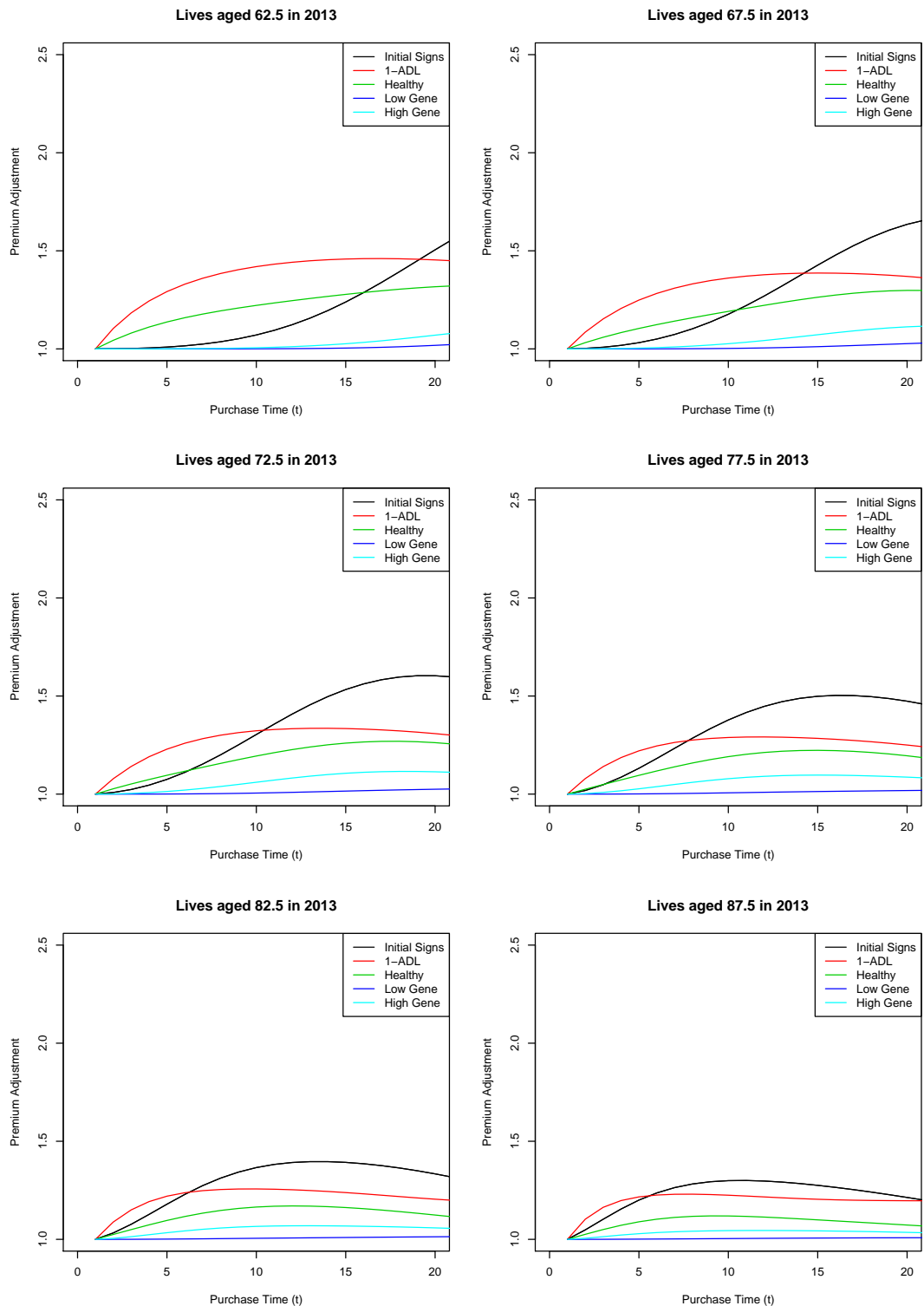
Figure 4.9: Repricing adjustments made to new business regular premiums through the emerging information from claims history, when a single source of adverse selection is present: "Healthy" — lives in the Healthy state don't buy; "Low Gene" — lives with AOPE genotypes $\varepsilon 2\varepsilon 2$ and $\varepsilon 3\varepsilon 3$ don't buy; "High Gene" — lives with AOPE genotypes $\varepsilon 3\varepsilon 4$ and $\varepsilon 4\varepsilon 4$ buy at an increased rate; "Initial Signs" — lives with initial signs of dementia buy at an increased rate; "1-ADL" — lives with 1-ADL functional disability type buy at an increased rate.
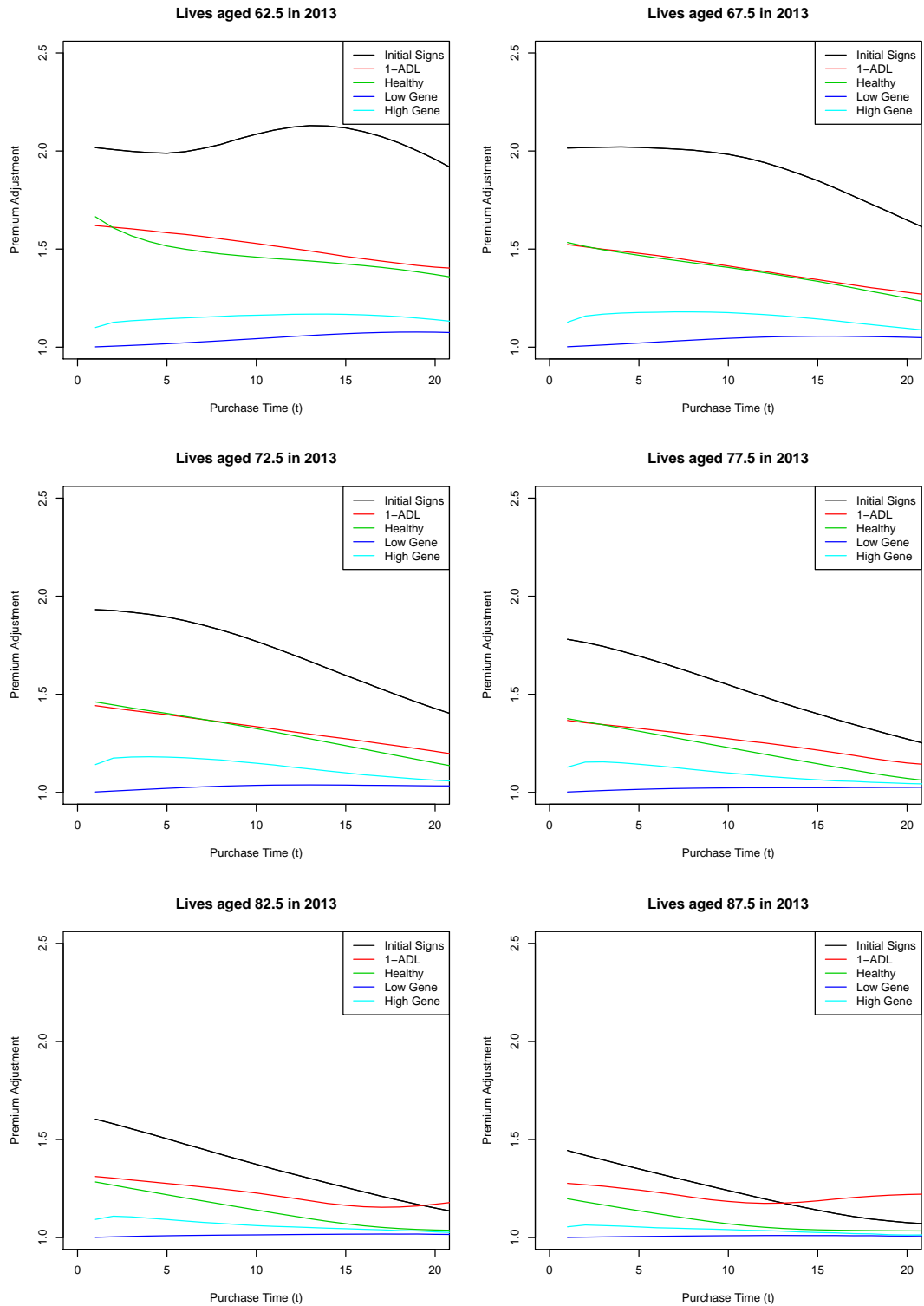
Figure 4.10: Adjustments which would be made to new business regular premiums if the insurer knew everything relevant about the customer at the point of sale, when a single source of adverse selection is present: "Healthy" — lives in the Healthy state don't buy; "Low Gene" — lives with AOPE genotypes $\varepsilon2\varepsilon2$ and $\varepsilon3\varepsilon3$ don't buy; "High Gene" — lives with AOPE genotypes $\varepsilon3\varepsilon4$ and $\varepsilon4\varepsilon4$ buy at an increased rate; "Initial Signs" — lives with initial signs of dementia buy at an increased rate; "1-ADL" — lives with 1-ADL functional disability type buy at an increased rate.

- Lives receiving in-home care will be given either $H = 1$ or $H = 2$ hours of care per day, corresponding to the results of Jones (2006); Forder and Fernández (2009) and the assumption in Nuttall et al. (1994)'s moderate needs estimate respectively. Costs of care are as outlined in Table 4.4.

- Testing will be performed at a rate of 0.08916 *per annum*.

- The force of inflation will be 0.02 or 0.04 *per annum*, while the force of interest is 0.05.

- An individual's care liability will be capped by the government at £75,000, adjusted for 4 years' inflation.

- The insurance contract will indemnify the policyholder up to £200,000 with no limit on annual payouts.

- The relevancy function in Equation (3.57) will be parameterised using $\beta = -0.01$ and $\gamma = -0.25$.

- The number of lives of age x and sex ς is as in Table 4.8 and denoted by $N(x, \varsigma,)$.

In our modelling we assume adverse purchase rates for males and females are equal. However this is not necessarily how it would be borne out in practice as the product represents a different value to each sex — females may be expected to have a higher purchase rate as the product represents greater value to them.

We calculate the cost of adverse selection for all business sold in underwriting class k during calendar year c by realising the future losses the business will incur when adverse selection is present, and discounting these to start of year c. We assume the premiums are calculated using the claims history to apply a repricing adjustment to the base premiums, the latter calculated assuming no adverse selection. By expressing the loss as a percentage of premium income, we show how much further the premiums need to be increased, in order to reach the actuarially fair premiums that were intended. Benefit outgo and premium income arising from business sold to lives aged x at 1st January, 2013, are weighted by the number of lives in that age group. This can be expressed as,

$$\text{Adverse Selection Cost} = \frac{E(\text{PV Benefits-PV Premium}|\text{Adverse Selection})}{E(\text{PV Premium}|\text{Adverse Selection})}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}, \varsigma \in \{M,F\}, g \in \mathcal{G}, i \in \mathcal{S}, j \in \mathcal{U}_k, \vartheta \in \mathbb{B}} N(x,\varsigma) \int_c^{c+1} p_{x,\varsigma,g}^i \, {}_t\bar{p}_{x,\varsigma,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,\varsigma,g}^{0\vartheta j,1\vartheta j} \left(A_{x,\varsigma,g,t:c}^{j,P} - \overline{\Pi}_{x,t}^k \, a_{x,\varsigma,g,t:c}^{j,P} \right) dt}{\sum\limits_{x \in \mathcal{X}, \varsigma \in \{M,F\}, g \in \mathcal{G}, i \in \mathcal{S}, j \in \mathcal{U}_k, \vartheta \in \mathbb{B}} N(x,\varsigma) \int_c^{c+1} p_{x,\varsigma,g}^i \, {}_t\bar{p}_{x,\varsigma,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,\varsigma,g}^{0\vartheta j,1\vartheta j} \, \overline{\Pi}_{x,t}^k \, a_{x,\varsigma,g,t:c}^{j,P} \, dt}, \quad (4.5)$$

where $\mathcal{X} = \{62.5, 67.5, 72.5, 77.5, 82.5, 87.5\}$ is the set of midpoints of each of our age groups.

We start by considering adverse selection costs when lives with the initial signs of dementia and lives with 1-ADL buy at an increased rate without any underwriting. The progression of these costs are shown in Figure 4.11 for regular and single premium on unisex and gender specific bases with $H = 1$ and $\nu = 0.04$.

The patterns shown are much the same regardless of the premium basis: very high costs attributable to business sold at the start of the market, while policies sold after 20 years since market set-up have negative costs. These negative costs reflect the repricing process increasing premiums too much — referring back to Figure 4.10, we see the premium adjustments which should be made gradually decrease, which is due to the base premium accommodating a greater proportion of higher risk lives without adverse selection. Losses are noticed more rapidly when the purchase rate for adverse selectors is highest, and despite being the highest cost initially, the smallest market with high adverse purchase rate responds quickest and becomes the most negative. The regular premium version of the contract has approximately 30% higher costs initially and these costs become more negative when policies are being charged too much.

There are a number of uncertain parameters involved in our model. To assess how the choice of value of these parameters might influence results, we present some analysis of the sensitivity.

The sensitivity of adverse selection costs over time to the inflation rate parameter can be seen in Figure 4.12 to be very low — the shape of the emergence over time and the size are very similar. Since the highest costs are at the start of the market, before the insurer has been able to adjust their correct their pricing assumptions, when the inflation rate is higher (and consequently the real discount rate lower) the smaller cost incurring/profit making business in the latest years is given more weight making the overall adverse selection cost, as a percentage of premium income, lower.

Figure 4.12 also shows the sensitivity to the number of hours of in-home care provision, H. Increasing H from 1 to 2 decreases the adverse selection cost. This may seem counter intuitive since we have lives at risk of claiming in-home care provision buying at an increased rate. However, as noted above, the biggest part of adverse selection cost is from lives observing the initial signs of dementia. These lives are also exposed to the parameter H, pre-institutionalisation but once they have been institutionalised, their care costs are no longer based on H. Consequently, although overall benefit costs are higher when $H = 2$, the increased premium rate offsets some of the adverse selection costs. Since the shape of the development of the costs does not change markedly, only the scale, we continue in this section with $H = 1$.

Now when we include adverse selection from lives who have had a genetic test (shown in Figure 4.13), the effect is two-fold: initially the costs are lower with genetic adverse selection due to lives with high-risk genes buying insurance earlier than they
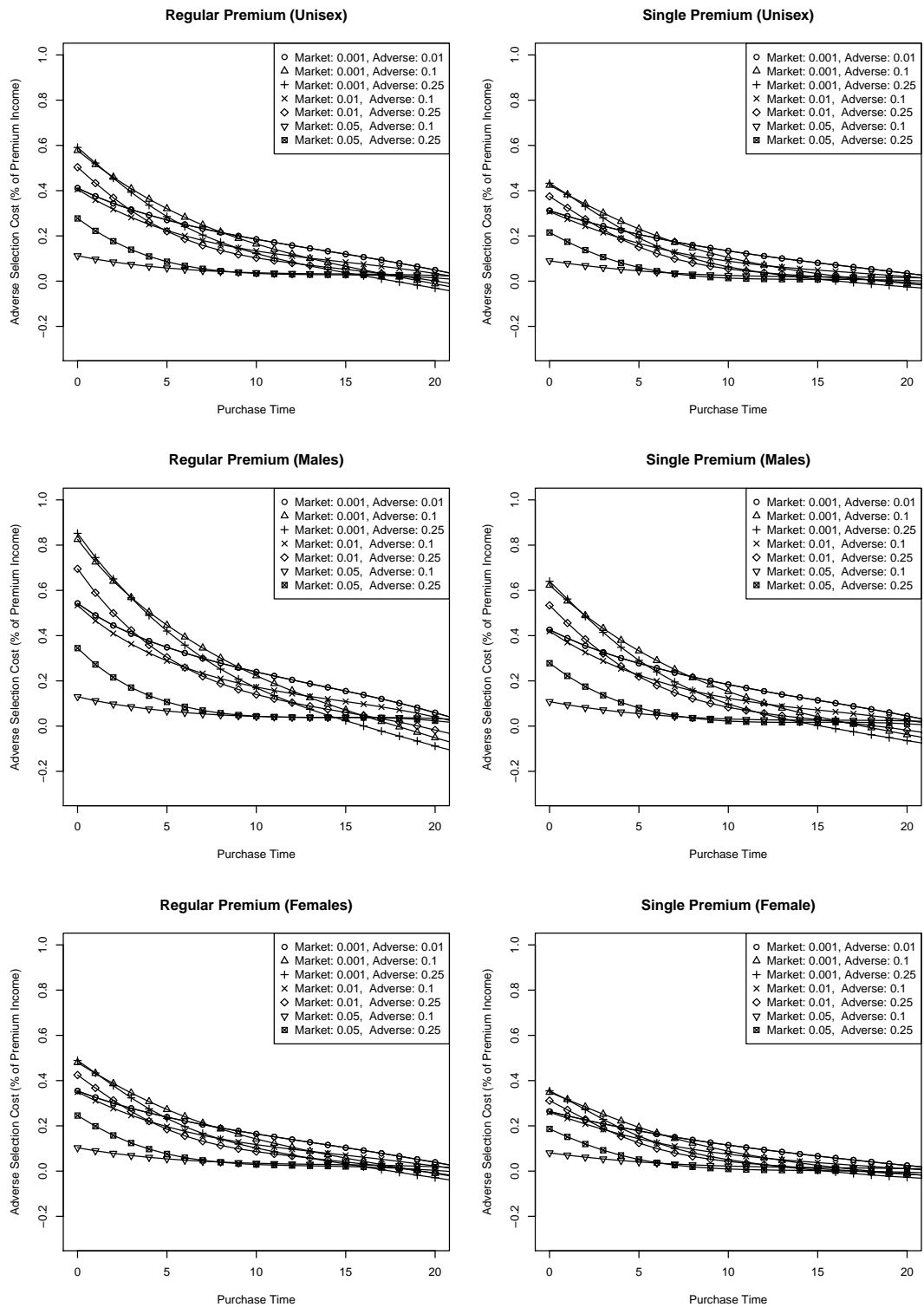
Figure 4.11: Progression of adverse selection cost when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate with $H = 1$, $\nu = 0.04$.
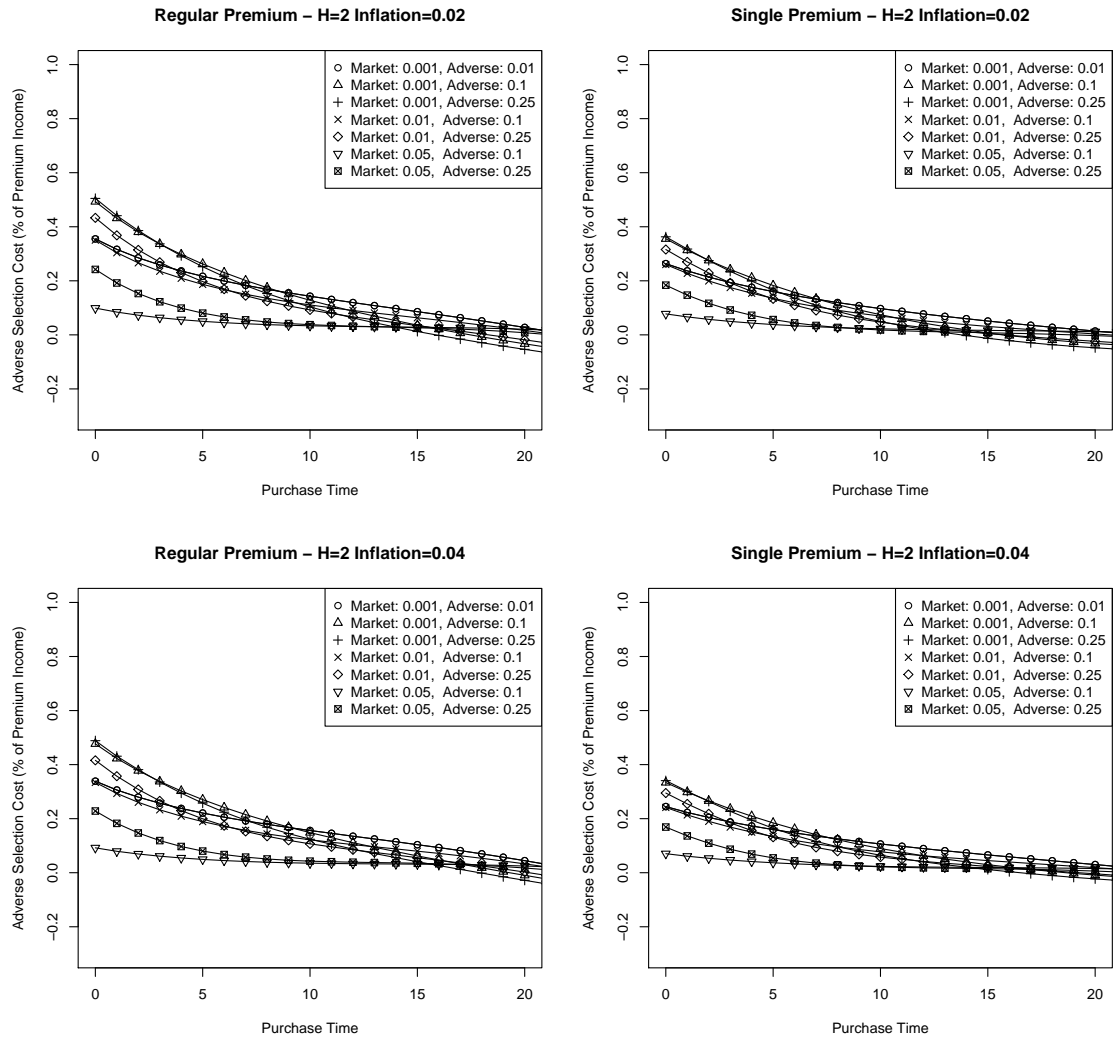
Figure 4.12: Sensitivity of adverse selection costs to number of hours of care provision and the force of inflation when lives with the initial signs of dementia and lives with 1-ADL buy insurance at an increased rate and premiums are charged on a unisex basis.

would when they move to a state where the difference between their benefit costs and premium income is greater; at later purchase times, the adverse selection cost is higher with genetic adverse selection because losses due to genetic adverse selection are realised slowly (see Figure 4.9) so premiums have not been changed sufficiently. In the large market, this creates a local minimum and subsequently a local maximum for males and where premiums are unisex.

If the market fails to attract lives in the Healthy state but lives with the initial signs or 1-ADL buy at an increased rate there will be little difference between the market sizes (see Figure 4.14) since only the relatively small proportion of lives with an IADL buy at the standard market rate. The effect of genetic testing in this case is minimal since the only lives in the IADL state will change buying behaviour due to test results.

The scenarios above have assumed there is a restriction on underwriting. If we permit insurers to underwrite based on functional ability, as would be the case in the UK, there would be no reason for lives with 1-ADL to purchase insurance at an increased rate, merely because of their functional disability.

Since our model is Markov, a life's health history (beyond its current state) tells nothing of it's future health. Therefore when a life has been underwritten into a high risk class, if the life subsequently recovers they could lapse their policy and buy another written in a lower risk class with a correspondingly lower premium. This is a consideration that would be necessary only when charging regular premiums, similar to the problem mentioned in Section 4.4.

The results of introducing underwriting are shown in Figure 4.15 when lives with the initial signs of dementia buy insurance at an increased rate and change buying behaviour after the result of a genetic test. Despite eliminating functional ability as a source of adverse selection, the costs as a percentage of premium income have increased when we compare to Figure 4.13, particularly for females and the smaller market sizes. This should be expected because the biggest part of adverse selection cost is from the insurer being unable to discern between lives with and without the initial signs of dementia. This still remains, while lives in the 'low risk' underwriting class pay a smaller premium than without underwriting. Hence, the nominal amount of loss has decreased, but less than the nominal amount of premium income has. Additionally, the local maximum observed upon introducing genetic adverse selection is more pronounced and exists for the other market sizes. Losses due to adverse selection from 1-ADL are realised relatively rapidly hence increased premiums can be used to cover losses from high purchase rates from lives with initial signs of dementia which are realised more slowly.
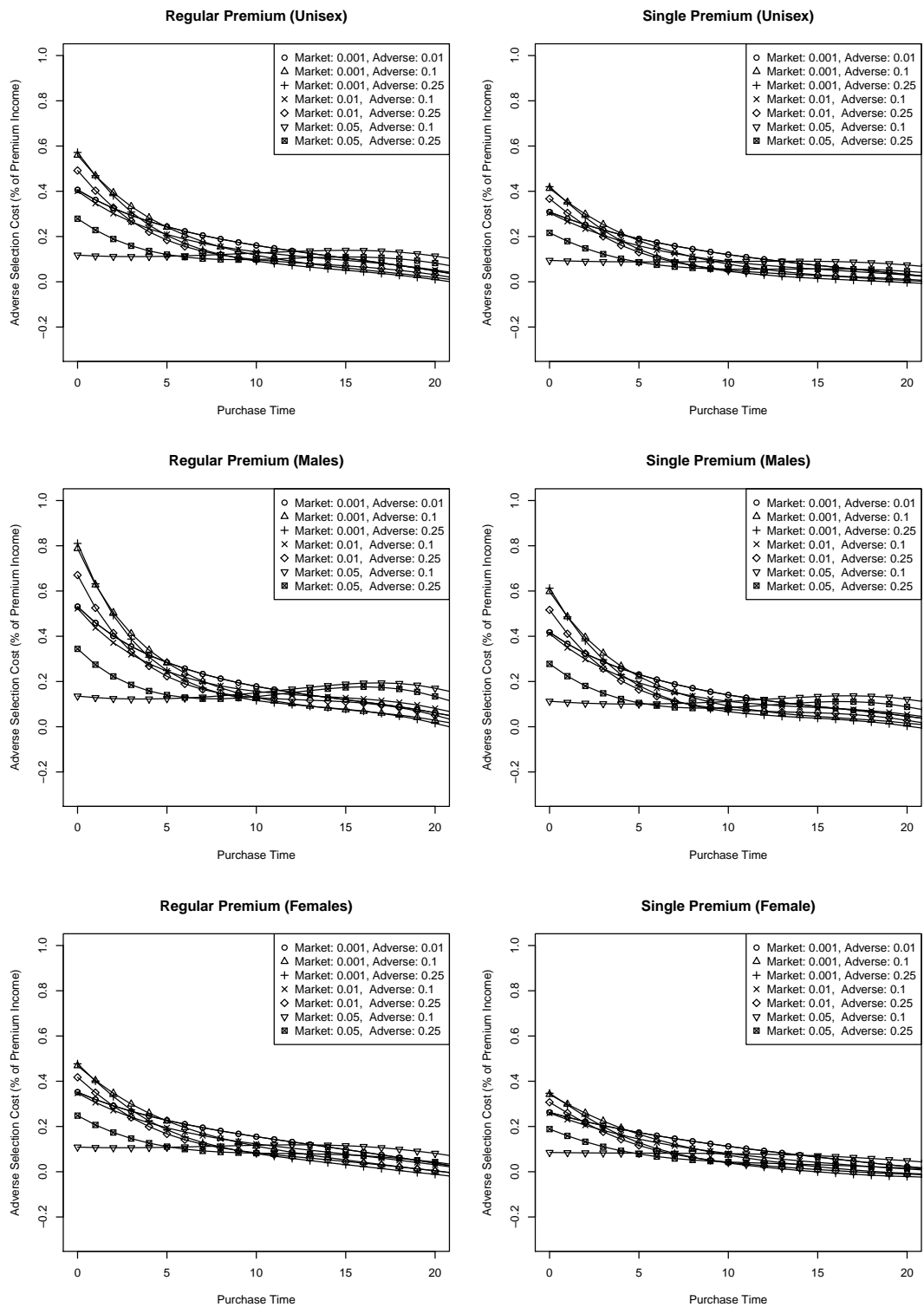
Figure 4.13: Progression of adverse selection cost when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate and lives change buying behaviour after having a genetic test, with $H = 1$, $\nu = 0.04$.
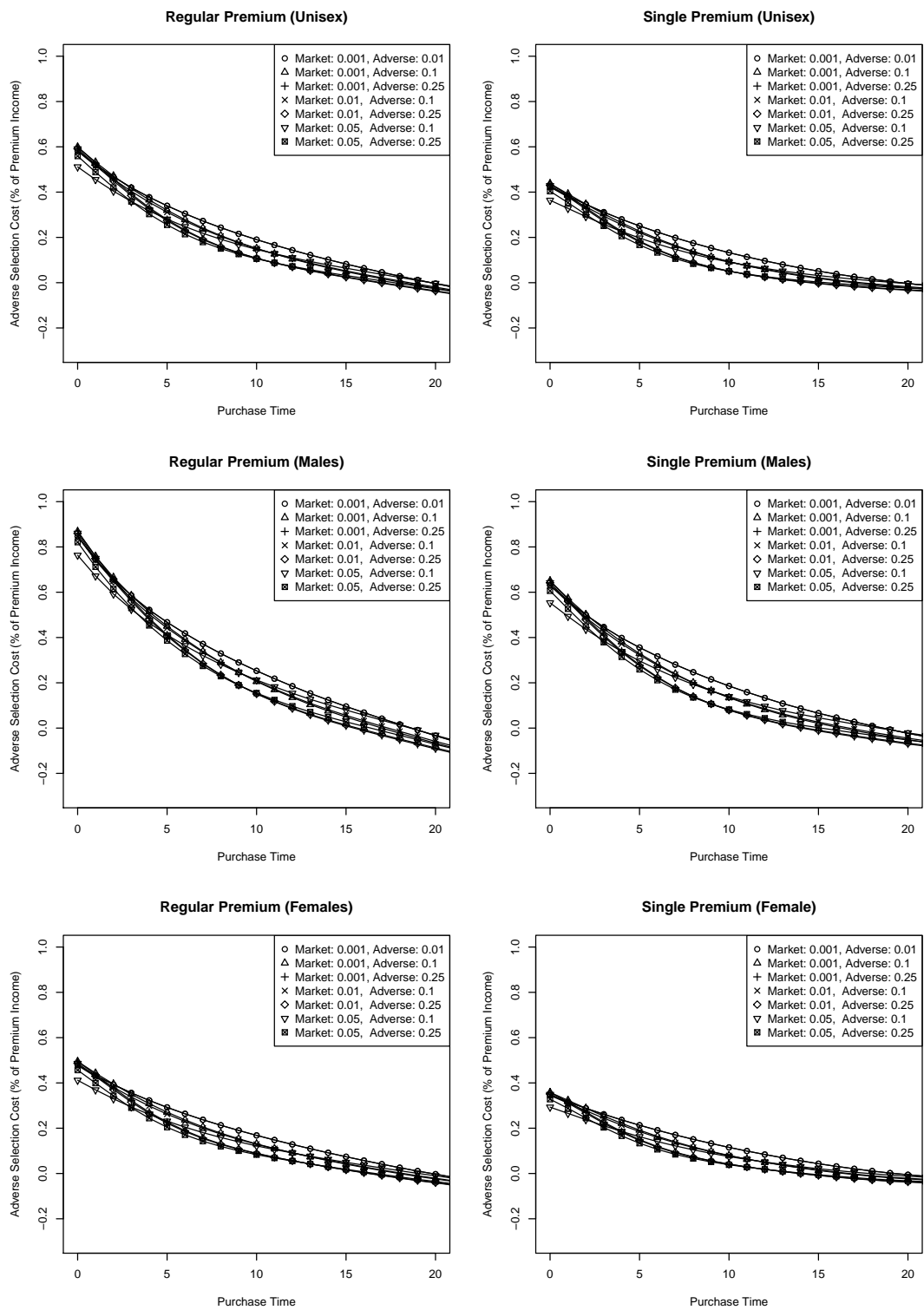
Figure 4.14: Progression of adverse selection cost when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate and lives change buying behaviour after having a genetic test, while healthy lives do not buy insurance regardless of genotype, with $H = 1$, $\nu = 0.04$.

148
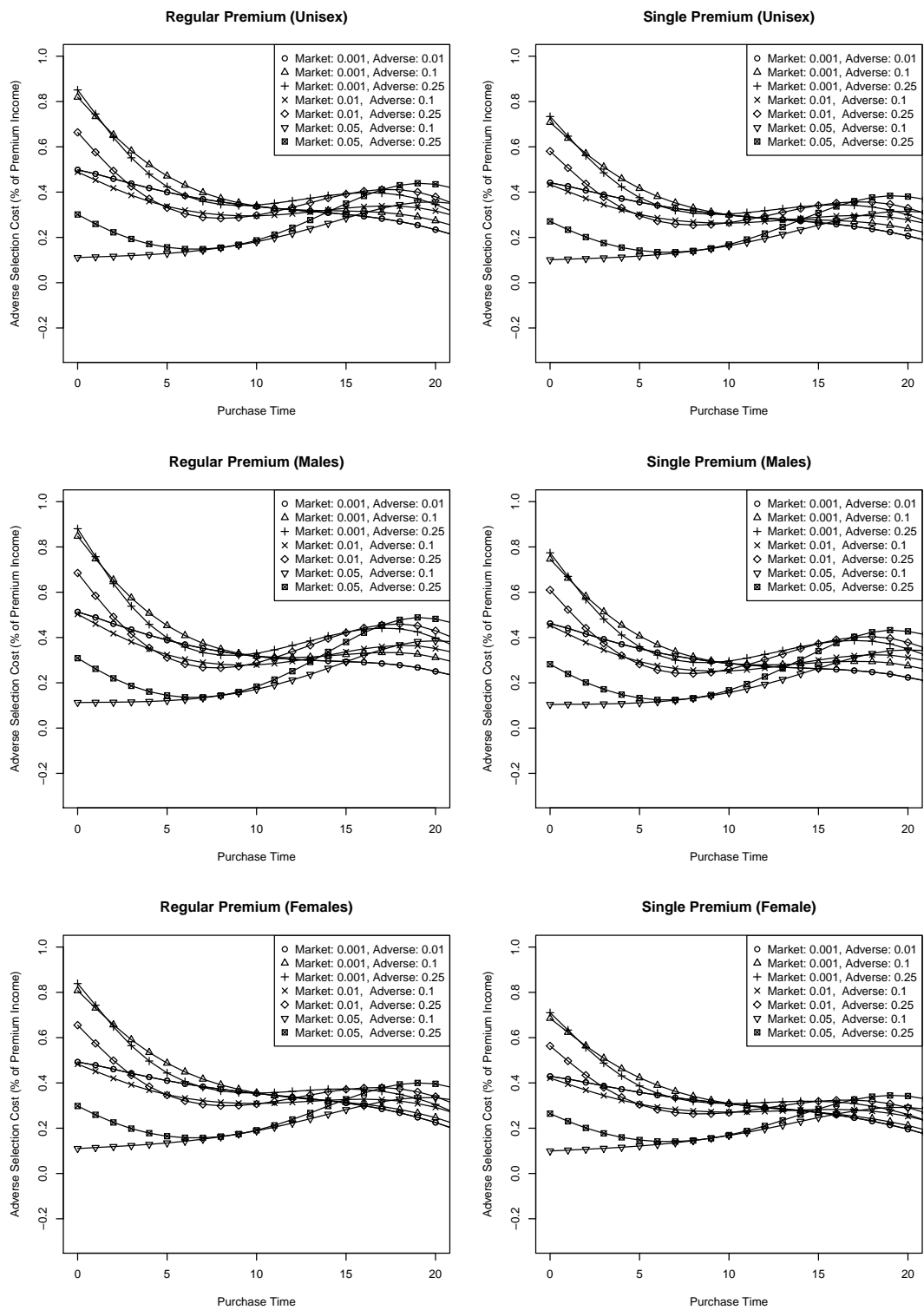
Figure 4.15: Progression of adverse selection cost when lives with 1-ADL are written into a separate class while lives with the initial signs of dementia buy insurance at an increased rate and lives change buying behaviour after having a genetic test, with $H = 1$, $\nu = 0.04$.

4.5.3 Varying the Caps

We have used a cap on care costs before the government takes on the liability assuming £75,000 in 2017, as per the Westminster government's stated intentions. If this does not get approval of parliament, there will be no cap on the liability faced by the individual. Moreover, we have not considered the situation in Scotland, where care costs are, intended to be met by the Scottish government. We examine these extremes of policy here.

First we consider the Scottish government's policy and assume that their payment is sufficient to meet the care costs. The adverse selection costs (shown in Figure 4.16) as a percentage of premium are substantially larger than those where the insurance policy also pays for care costs. In this situation, the role of the insurance policy is to meet the hotel costs, a benefit only associated with dementia. Correspondingly, benefit sizes are smaller and the probability of reaching a claim is reduced, hence the premiums are also reduced. This smaller premium base, as we have seen above (in the scenario with underwriting), creates large adverse selection costs.

Next we look at adverse selection costs if the U.K. government does not implement their proposed cap. These are shown in Figure 4.17. Comparing these with the equivalent scenario with the cap, in Figure 4.11, we see very little impact from the removal of the cap on care costs.

To investigate this further, we amended our simulation model to estimate the distribution of total care individuals face over their lifetime. For each sex, genotype and health state, we simulated the future life histories and the associated care costs of 1,000,000 lives aged $x \in \mathcal{X}$. Where the total care costs for lives was non-zero, these were separated into 'bins' of size £1,000 in terms of 1 January, 2013's purchasing power. A further bin existed for lives who never had any care expenditure. Summing the number of lives in each bin over genotypes $g \in \mathcal{G}$ and states $j \in \mathcal{S}$, with weighting equal to the probability for the initial distribution, $p_{x,\varsigma,g}^{j}$ (calculated in Section 4.2.2), allows us to estimate a distribution for a discretised care cost, for a particular age group x and sex ς.

The nature of adverse selection means that in our pool of insurance business, the mix of lives who buy insurance does not match the mix of lives in the general population — there will be a higher proportion of lives who require care. As a proxy for the mix of insured business, we consider the distribution of care costs, conditional on the care cost being non-zero. Denote the number of lives aged x, of sex ς, with genotype g, and in state j at 1st January, 2013, whose simulated care costs were between $1000i$ and $1000(i+1)$, by $b_{x,\varsigma,g}^{j}(i)$. For lives aged x, of sex ς, we estimate the probability mass function of care costs, conditional on the cost of care being non-zero,

Regular Premium (Unisex)

Single Premium (Unisex)

Regular Premium (Males)

Single Premium (Males)

Regular Premium (Females)

Single Premium (Female)
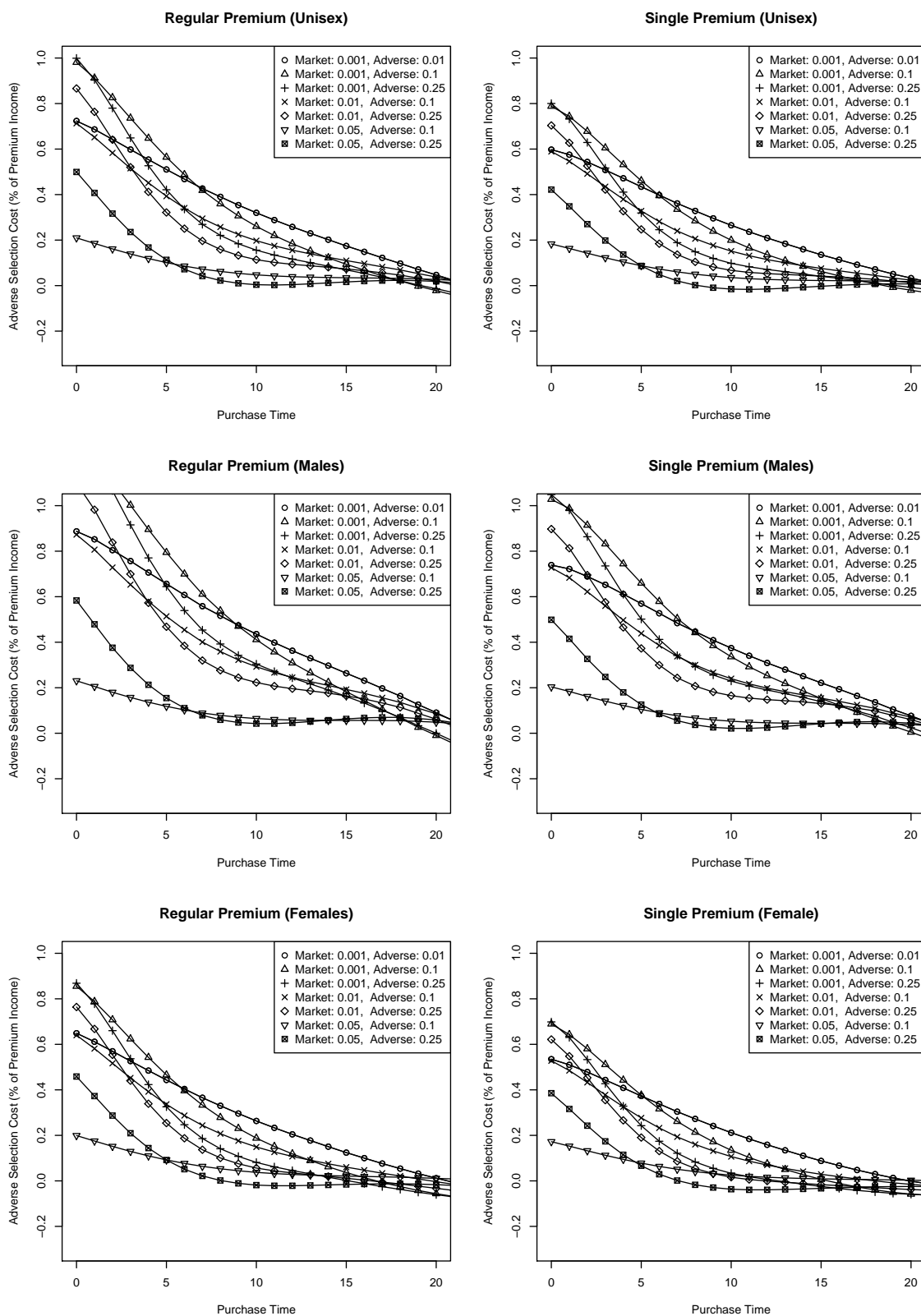
Figure 4.16: Progression of adverse selection cost when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate with $H = 1$, $\nu = 0.04$. Insurance pays for hotel costs only.
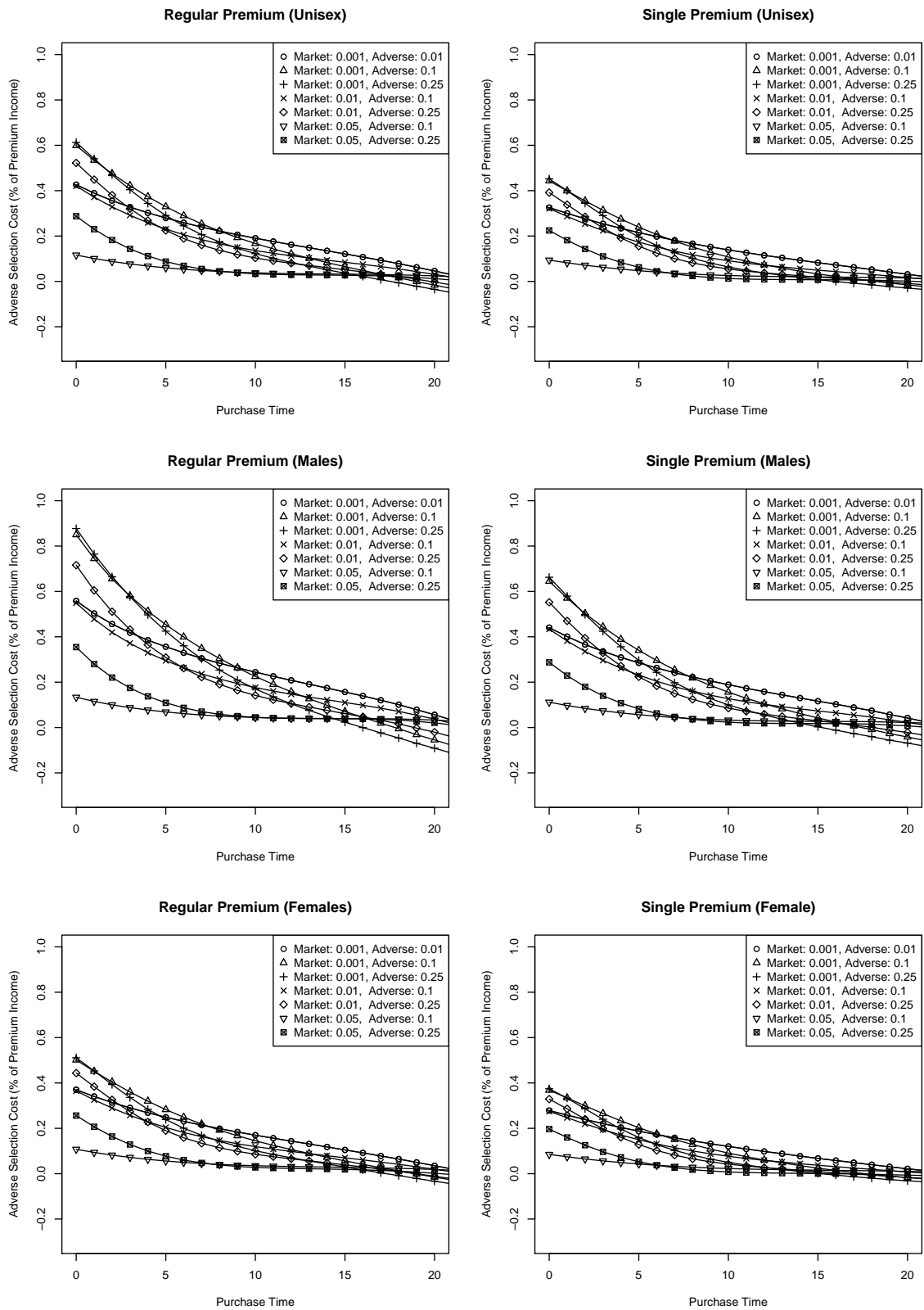
Figure 4.17: Progression of adverse selection cost when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate with $H = 1$, $\nu = 0.04$. The government does not cap care liability.
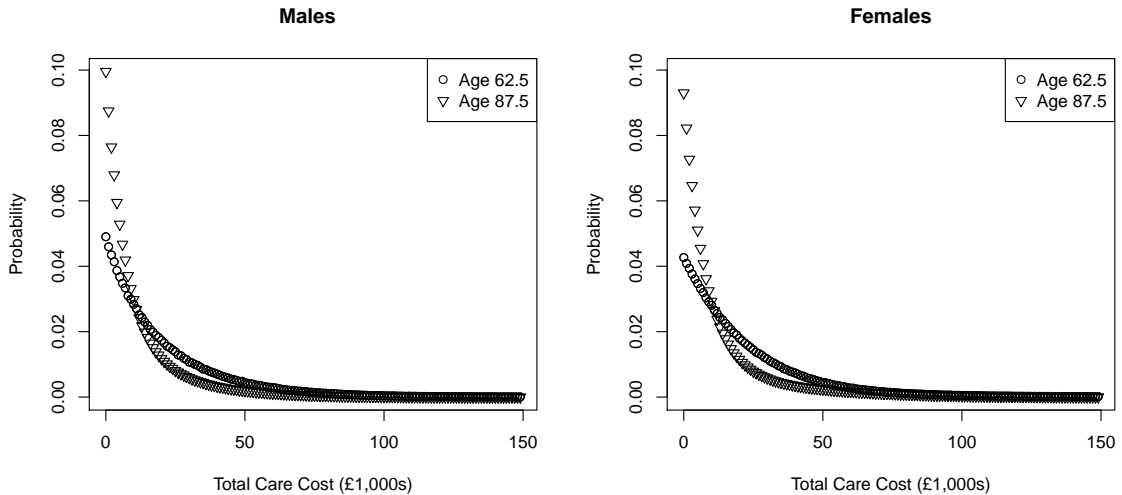
Figure 4.18: Probability density functions of care costs, conditional on the cost of care being non-zero, separated by age and sex.

as,

$$P\left(1000i < \text{Care cost} \leq 1000(i+1)|\text{Care cost} > 0\right) = \frac{\sum\limits_{g \in \mathcal{G}, j \in \mathcal{S}} p^j_{x,\varsigma,g}\, b^j_{x,\varsigma,g}(i)}{\sum\limits_{g \in \mathcal{G}, j \in \mathcal{S}, k \in \mathbb{N}} p^j_{x,\varsigma,g}\, b^j_{x,\varsigma,g}(k)}.$$

$$(4.6)$$

For ages 62.5 and 87.5, this is shown in Figure 4.18.

The adverse selection we have considered does not impact upon the duration of care requirements. We therefore suggest the distribution of care costs would not differ greatly between what is anticipated in pricing and what is experienced. Since only a small proportion of lives are affected by the removal of the cap (see Table 4.10), where a difference in this distribution were to arise, the associated loss would not be large when spread over a portfolio of business. This explains why the changes in costs of adverse selection were so small when the government cap was removed.

4.6 Cross-subsidy in Unisex Premiums

As noted in Section 4.4, premiums for females are substantially higher than those for males when using sex as as a pricing factor. Hence, under a unisex format where the costs are spread over both sexes, males will be paying more than they otherwise would.

Define the cross-subsidy to be the excess of premium income from one group (call this group A) over the benefits paid to them used to cover losses on business from

Table 4.10: Conditional probabilities of reaching the U.K. government's proposed care cap of £75,000, adjusted for 4 years' inflation, assuming a force of inflation of $\nu = 0.04$ *per annum.*

Sex	Age	P(Care costs=0)	P(Reach Proposed Cap—Require Care)
Male	62.5	0.5574	0.0516
	67.5	0.5338	0.0521
	72.5	0.5362	0.0479
	77.5	0.5402	0.0474
	82.5	0.5446	0.0374
	87.5	0.5430	0.0190
Female	62.5	0.3351	0.0617
	67.5	0.3109	0.0689
	72.5	0.3045	0.0734
	77.5	0.2986	0.0786
	82.5	0.2961	0.0706
	87.5	0.3033	0.0408

another group (call this group B):

$$
\text{Cross subsidy} = \begin{cases} \text{Profit from } A, & \text{if } 0 < \text{Profit from } A \leq \text{Loss from } B \\ \text{Loss from } B, & \text{if } 0 < \text{Loss from } B < \text{Profit from } A \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}
$$

For the existence of a cross-subsidy, it is therefore necessary for a simultaneous occurrence of both profit from one group of lives and losses from the other.

When the mix of lives matches what is priced for, this will create a cross-subsidy which is measurable simply by comparing the premium rates in Figures 4.3 to 4.8. However, in the context of adverse selection, this cross subsidy may disappear if premium income from males is insufficient to cover the benefits males receive or if the premium adjustment is such that females' benefits are fully covered by females' premiums. In this section we analyse the premium income and benefit payments at the level of particular sex to quantify the cross-subsidy in the unisex premiums. In this case group A are the male lives and group B are the female lives.

We can calculate the profit from males aged x in 2013, who purchased insurance in calendar year c, which we denote by $\overline{A}(x, c)$ and $A(x, c)$, for regular and single

premium versions respectively, as

$$\overline{A}(x,c)$$

$$= \sum_{g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,M) \int_c^{c+1} p_{x,M,g}^i \, _t\tilde{p}_{x,M,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,M,g}^{0\vartheta j,1\vartheta j} \left(\overline{\Pi}_{x,t}^k \, a_{x,M,g,t:c}^{j,P} - A_{x,M,g,t:c}^{j,P} \right) dt,$$

$$(4.8)$$

and

$$A(x,c) = \sum_{g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,M) \int_c^{c+1} p_{x,M,g}^i \, _t\tilde{p}_{x,M,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,M,g}^{0\vartheta j,1\vartheta j} \left(\Pi_{x,t}^k - A_{x,M,g,t:c}^{j,P} \right) dt.$$

$$(4.9)$$

Similarly, we calculate the loss from females aged x in 2013, who purchased insurance in calendar year c, which we denote by $\overline{B}(x,t)$ and $B(x,t)$, as

$$\overline{B}(x,c)$$

$$= \sum_{g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,F) \int_c^{c+1} p_{x,F,g}^i \, _t\tilde{p}_{x,F,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,F,g}^{0\vartheta j,1\vartheta j} \left(A_{x,F,g,t:c}^{j,P} - \overline{\Pi}_{x,t}^k \, a_{x,F,g,t:c}^{j,P} \right) dt,$$

$$(4.10)$$

and

$$B(x,c) = \sum_{g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,F) \int_c^{c+1} p_{x,F,g}^i \, _t\tilde{p}_{x,F,g}^{i,j} \, \tilde{\mu}_{x,t,F,g}^{j,INS} \left(A_{x,F,g,t:c}^{j,P} - \Pi_{x,t}^k \right) dt,$$

$$(4.11)$$

for regular and single premium versions respectively.

Let the cross-subsidy across all business sold in calendar year c, expressed as a percentage of male premium, be denoted by $\overline{\chi}(c)$ and $\chi(c)$ and calculated as

$$\overline{\chi}(c) = \frac{\min\left[\max\left(0, \sum\limits_{x\in\mathcal{X}} \overline{A}(x,c) \right), \max\left(0, \sum\limits_{x\in\mathcal{X}} \overline{B}(x,c) \right) \right]}{\sum\limits_{x\in\mathcal{X},g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,M) \int_c^{c+1} p_{x,M,g}^i \, _t\tilde{p}_{x,M,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,M,g}^{0\vartheta j,1\vartheta j} \, \overline{\Pi}_{x,c}^k \, a_{x,M,g,t:c}^{j,P} \, dt}, \quad (4.12)$$

and

$$\chi(c) = \frac{\min\left[\max\left(0, \sum\limits_{x\in\mathcal{X}} A(x,c) \right), \max\left(0, \sum\limits_{x\in\mathcal{X}} B(x,c) \right) \right]}{\sum\limits_{x\in\mathcal{X},g\in\mathcal{G},i\in\mathcal{S},j\in\mathcal{U}_k,\vartheta\in\mathbb{B}} N(x,M) \int_c^{c+1} p_{x,M,g}^i \, _t\tilde{p}_{x,M,g}^{00i,0\vartheta j} \, \tilde{\mu}_{x,t,M,g}^{0\vartheta j,1\vartheta j} \, \Pi_{x,c}^k \, dt}, \quad (4.13)$$

for regular and single premium versions respectively.

We analyse the cross-subsidies in the the same scenarios as above in Section 4.5.
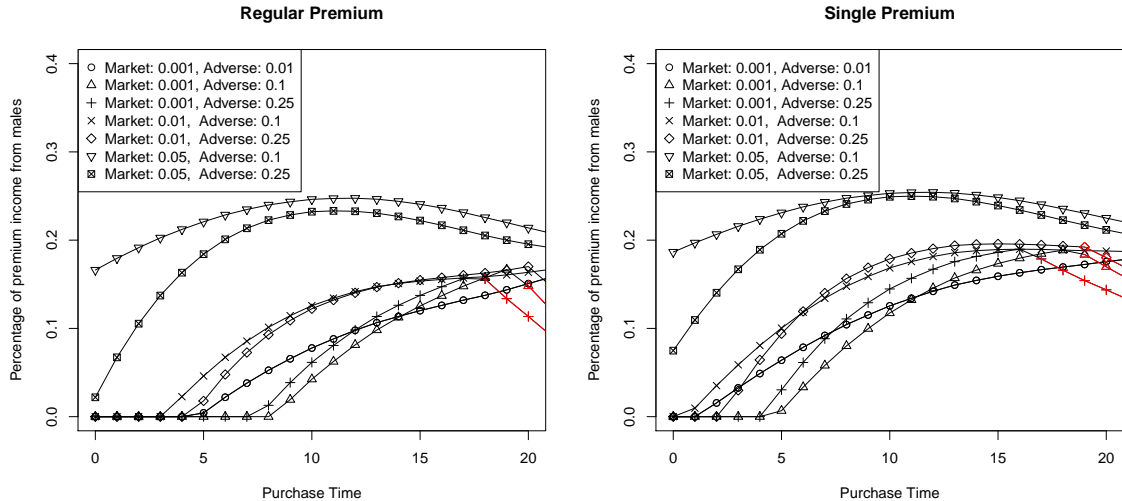
Figure 4.19: Progression of the cross-subsidy from males in unisex premiums when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate, with $H = 1$, $\nu = 0.04$. Black indicates the cross-subsidy is the profit on males, while red indicates that it is the loss from females.

Figure 4.19 shows the progression of the cross-subsidy when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate. There is an inverse relationship between cost of adverse selection and the cross-subsidy — while adverse selection cost is high, it is unlikely there is a profit being made from males and when adverse selection cost is negative, it is possible that there is no loss on female business. Consequently, the large market has the greatest cross-subsidy and is non-zero throughout. Smaller market sizes make a loss on both male and female business initially so have no cross subsidy. This lasts longer under the regular premium product than the single premium version.

When we introduce genetic testing to this scenario (see Figure 4.20), the shape shows some distortion in a similar fashion to the adverse selection cost but with local minima in place of local maxima.

If healthy lives do not buy insurance but lives with the initial signs of dementia buy insurance at an increased rate and lives alter buying behaviour after a genetic test (see Figure 4.21) then a loss is made on males in all market sizes, lasting between 7 and 9 years in the regular premium version and 4 and 6 years for the single premium version. At which point the cross subsidy begins increasing until a total profit is made and starts to decrease as the loss on females decreases due to the premium adjustment being made.

In Figure 4.22 we have lives with 1-ADL being underwritten into a separate class of business. Lives with the initial signs of dementia buy insurance at an increased rate and lives alter buying behaviour after a genetic test. In this case, a cross subsidy
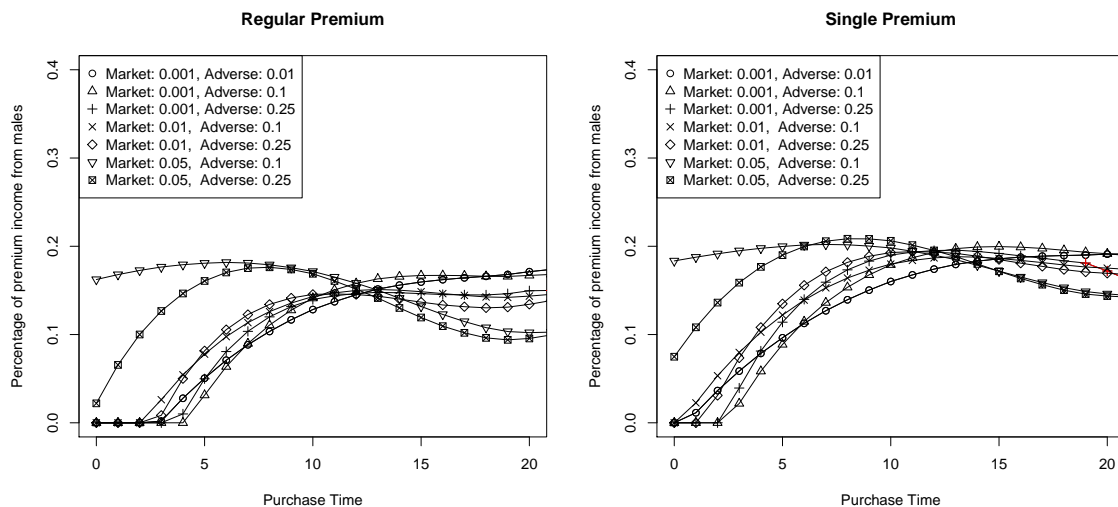
Figure 4.20: Progression of the cross-subsidy from males in unisex premiums when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate and lives change buying behaviour after having a genetic test, with $H = 1$, $\nu = 0.04$. Black indicates the cross-subsidy is the profit on males, while red indicates that it is the loss from females.
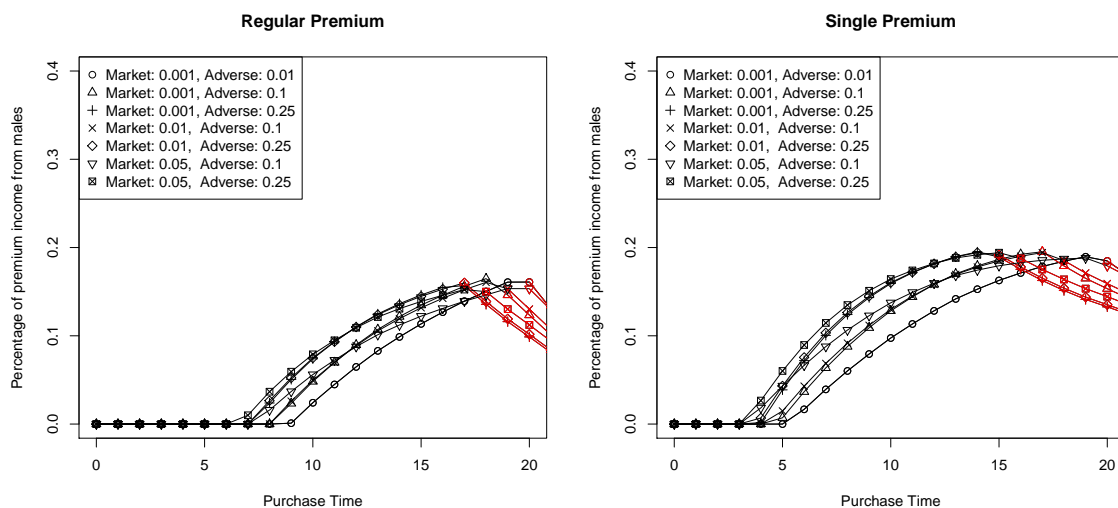
Figure 4.21: Progression of the cross-subsidy from males in unisex premiums when lives with the initial signs of dementia or 1-ADL buy insurance at an increased rate and healthy lives do not buy insurance and lives change buying behaviour after having a genetic test, with $H = 1$, $\nu = 0.04$. Black indicates the cross-subsidy is the profit on males, while red indicates that it is the loss from females.
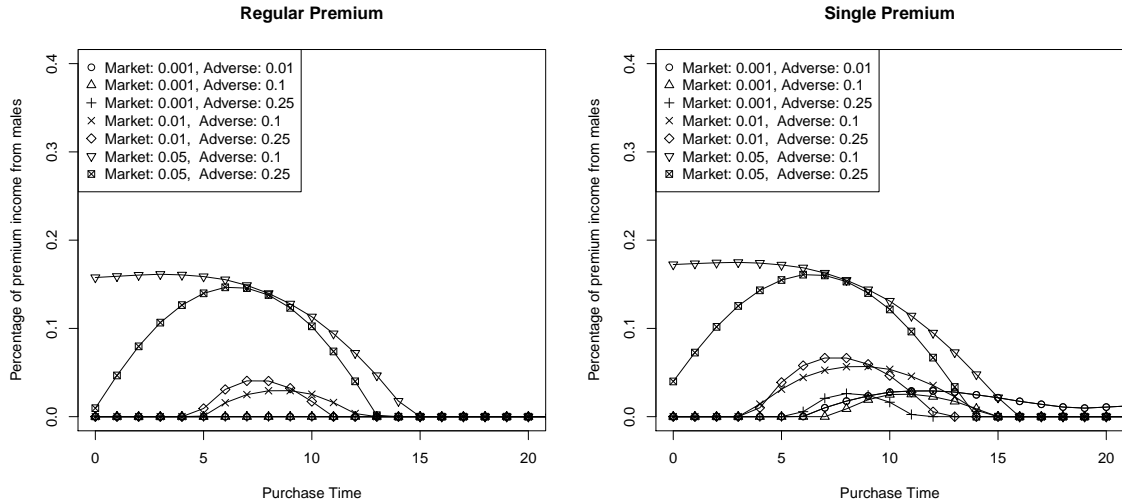
Figure 4.22: Progression of the cross-subsidy from males in unisex premiums when lives with 1-ADL are written into a separate class while lives with the initial signs of dementia buy insurance at an increased rate and lives change buying behaviour after having a genetic test, with $H = 1$, $\nu = 0.04$. Black indicates the cross-subsidy is the profit on males, while red indicates that it is the loss from females.

initially exists in the large market but becomes zero around year 15. Other market sizes see a cross-subsidy once premiums have been adjusted: in the single premium version, cross-subsidy exists for all cases but is short-lived for all but the lowest rate of adverse selection in the small market; under a regular premium, cross-subsidy only exists in the smallest market in the final years, and lasts for a short time in market with purchase rate 0.01.

4.7 Conclusions

Renewed interest in long-term care from the government, and coverage in the press highlighting the need for reform, has made the prospect of creating a new LTC market in the U.K. a possibility. Adverse selection has been suggested as a potential barrier to the success of such a market. This adverse selection could come from lives who have had their APOE genotype tested, have the initial signs of dementia, are healthy, or depending on underwriting practice, have some functional disability.

Our dynamic pricing strategy allows insurers to correct their premiums. We have seen that losses from dementia are realised slower than losses from functional ability, but it is dementia where the the greatest adverse selection costs arise.

These costs vary greatly over time. The reason for this is twofold: the insurer adjusts its premium rates to factor in its experience; and the base premiums assume increasing proportions of the high risk lives — who are adversely selecting. In the

early years of the market, where the majority of the modelled lives purchase their contract, they can reach significantly large levels. In the context of large premiums on contracts being sold to pensioners on tight budgets, any increase of premiums to pay for adverse selection would present challenges in marketability.

The adverse selection costs are insensitive to movements of inflation and to the level of benefits paid for functional disability. While an improvement in the size of the government's cap on care liability changes adverse selection costs significantly, removing the cap made very little impact. We found this was due to there being a small percentage of lives benefiting from the introduction of such a cap. How well this meets the needs of citizens is a matter for social policy researchers to consider and politicians to debate, however actuaries will be interested in the extent that it removes tail risk, which our results cannot say anything about.

Within unisex premiums for LTC, there is a high degree of cross-subsidy from males to pay for the higher benefit costs of females. LTC is not unique in such a cross-subsidy, a common example of a contract with one being motor insurance (where females subsidise males), but introducing a new product with cross-subsidy may make selling to men difficult. Whereas our pricing responds immediately to differences between the genders, an insurer might take somewhat longer.

Chapter 5

Conclusions and Further Work

The work we have done has been to analyse the costs of adverse selection in insurance markets where there are multiple causes of information asymmetry, including genetic information. In this chapter we conclude the work and highlight where further work may be done to explore the ideas covered in this work.

We remind the reader that the models we have presented were parameterised using data from various periods and populations. Moreover, the uncertainty of the estimated parameters has not been accounted for. While the compatibility of these sources and the uncertainty is not a concern for our work, these should be considered before our models are put to use for other purposes.

5.1 Conclusions

5.1.1 Breast Cancer

Chapter 2 was concerned with a polygenic model of BC, where a large number of common but low impact genetic variations independently alter the risk of developing BC. Working from epidemiology of newly discovered polygenes, we parameterised a model to estimate the distribution of polygene relative risk among women with and without a family history of breast or ovarian cancer. We found women with a family history have a significantly higher risk of developing breast cancer, making it a good proxy for genetic risk in the absence of genetic test results. We used our estimated relative risk distributions to estimate the distributions for the CI premiums chargeable to lives with and without family history and found that reasonable terms could be offered.

In the context of the European Court of Justice's sex discrimination ruling, we recalculated these as unisex premiums. We envisaged two potential underwriting scenarios for a family history of BC or OC among female relatives: only females are rated; or males and females given an equal rating. Ratings of the former differ slightly

due to the change in unrated premium from pattern of male claims being lighter than females at early ages but heavier at older ages. In the latter case, since there were approximately twice the lives over which the ratings are spread, the ratings were approximately halved, as was their standard deviations.

We expanded our model of CI into a market for the insurance product, with purchasing behaviour of individuals potentially changing after a genetic test. We calculated adverse selection costs under a range of cases, some of which had results which would be noticeable to an insurance company or a customer. From our scenarios, we were able to conclude that the greater part of the adverse selection cost is due to the polygene component, rather than the major genes, confirming the conclusion of Macdonald and McIvor (2009). Moreover, we found that lives with a low-risk polygenotype choosing not to buy the contract, could present a reasonable adverse selection cost.

5.1.2 Long-term Care Insurance

In Chapters 3 and 4 we were interested in a start-up market for LTC and how the costs of adverse selection varied over time. We first outlined and parameterised a model for old age health which included states for varying levels of functional ability and cognitive function. We added and parameterised a hypothetical, intermediate state of cognitive decline where we assumed the initial symptoms were observable to the individual but not yet diagnosable and hence, invisible to the insurer. This information asymmetry gave rise to a possibility of adverse selection from these states of health.

An insurer could be expected to reprice when it learns that the mix of lives buying the product differs from what was assumed. We incorporated a novel dynamic repricing methodology into our model, which adjusts some base premium based on claims history.

When we applied this model to the U.K. market, we found that the sources of adverse selection differed in the rate at which losses were realised, and hence, how fast an insurer was able to adjust premiums to account for the adverse selection it was experiencing. When adverse selection came from lives with the initial signs of dementia, the rate premiums were adjusted was slowest compared to the other potential sources of adverse selection.

On its own and with an exaggerated rate of testing, genetic testing poses a moderate cost which is dwarfed when compared to the other potential sources of adverse selection. Since a large proportion of lives with the initial signs of dementia progress to large claims, they were also responsible for the largest adverse selection costs. When other sources of adverse selection were included, in addition to the initial signs of dementia, adverse selection costs were decreased because the selection was able to

occur earlier, when lives were of a slightly lower risk. Moreover, claims from lives with functional disability adversely selecting, allowed the repricing mechanism to increase premiums to cover some of the losses to be incurred from the dementia claims.

We investigated the effect of the U.K. government's proposed cap on an individual's care liability. At the level which they are proposing, only a small proportion of people will face enough care costs to benefit from it. Consequently, it has very little impact on the costs of adverse selection.

5.2 Further Work

In our BC polygene model, we assumed independence in both inheritance and how the polygene acts on BC risk. Without evidence to the contrary, this assumption allowed us to create a simple model which gives good guidance. If the number of polygenes turns out to be as large as we have assumed, there will likely be numerous loci close enough together in the genome that they are often inherited together. Moreover, there could be interactions between loci, such that the effect of particular combinations differs to what the same combination would cause in our assumption of independence.

Another simplifying assumption was made with regards to our hypothetical state for the initial signs of cognitive decline, in that we assumed equal transition intensity into and out of the state. How well this model fits reality would need to be considered and, when further understanding of the progression of dementia becomes available, a re-parameterisation of this state would be appropriate.

Our model of an LTC market is dynamic from the perspective of the insurance company, in that it reflects the repricing activity that could be made based on claims and income experience. However, we do not make any consideration for where a buying behaviour might change in response to the relative cheapness of the premium rates, beyond simple deterministic scenarios. Our CI market model worked in a similar fashion. These could be improved by applying some form of decision making methodology, such as utility theory or prospect theory or using an elastic model of demand with a suitably parameterised price elasticity based on solid empirical data.

Given the high costs from adverse selection and the uncertainty over purchasing patterns, it is likely that there is a large risk inherent in writing LTC contracts. To quantify this risk, the LTC simulation model could be amended to calculate the distributions of benefit payments. From these distributions and a suitable model of purchases, measures of risk including value at risk could be calculated which would aid in understanding the risk of the product and how it is influenced by adverse selection. This would be especially important in the context of a risk-based capital requirement regime such as Solvency II. Product designs which share the insurance risk with the policyholder such as with-profits or reviewable premiums could act to

reduce the capital requirements. How well these types of products meet the needs of the customer would need further research, in a similar approach to that taken by Guillén and Comas-Herrera (2012) in Spanish LTC products.

Alternatively, a product which serves as a life annuity, bought out of pensions savings and also provides LTC benefits, might reduce the adverse selection costs. Macdonald and Pritchard (2001) considered this mixing of an annuity and LTC and found the adverse selection costs became much more manageable. However, this is not a solution for today's elderly but may help the situation in the future.

A recent case in the Supreme Court of the United States has prevented the patenting of naturally occurring genes[1] in the U.S. This will allow increased competition in the market for genetic tests and potentially increase innovation and bring down the costs involved. Cheaper testing and future developments in genomic medicine will make it more likely that gene testing for the purposes of treating or preventing diseases will become routine, although such a scenario is not yet on the horizon. Increased awareness of risks brings with it a potential for adverse selection if this information is not shared with an insurance company. Empirical evidence of adverse selection related to genetic testing is limited and further work is necessary to make solid conclusions.

The direction of genetic research is moving towards analysing the function of risk genes, to understand how they cause predisposition to particular diseases. Such improved understanding might in turn lead to the further development of genomic medicine. Advances in genomic medicine have been slow to surface, but a recent review by Manolio et al. (2013) is optimistic about its future if more collaboration were to take place. Moreover, recent results from the second Cognitive Function and Ageing Study reported by Matthews et al. (2013), showed a significant decrease in prevalence of dementia. They interpreted this as providing evidence for a cohort effect in dementia prevalence. Actuaries will need to adapt their modelling in line with where the epidemiology leads, whether this could mean the use of multi-factorial models which include the additional factors of *e.g.* cigarette smoking or weight, or how developments lead to changing patterns of morbidity over time.

---

[1]Association for Molecular Pathology vs. Myriad Genetics, Inc. 569 U.S. 12-398 (2013)

Bibliography

C. J. Adams, C. Donnelly, and A. S. Macdonald. The impact of known breast cancer polygenes on critical illness insurance. *Scandinavian Actuarial Journal*, Online: 1–31, 2013.

O. Aguilera, A. F. Fernández, A. Muñoz, and M. F. Fraga. Epigenetics and environment: a complex relationship. *Journal of Applied Physiology*, 109:243–251, 2010.

G. A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, pages 488–500, 1970.

O. Akodu. *A stochastic & dynamic projection model of long-term care demand and cost in the United Kingdom*. PhD thesis, Heriot-Watt University, 2007.

K. Aktan-Collan, A. Haukkala, J.-P. Mecklin, A. Uutela, and H. Kääriäinen. Psychological consequences of predictive genetic testing for hereditary non-polyposis colorectal cancer (HNPCC): A prospective follow-up study. *International Journal of Cancer*, 93:4:608–611, 2001.

K. Aktan-Collan, A. Haukkala, and H. Kääriäinen. Life and health insurance behaviour of individuals having undergone a predictive genetic testing programme for hereditary non-polyposis colorectal cancer. *Public Health Genomics*, 4:4:219–224, 2002.

K. Alagiakrishnan, S. S. Gill, and A. Fagarasanu. Genetics and epigenetics of Alzheimer's disease. *Postgraduate Medical Journal*, 88:522–529, 2012.

T. d. S. Alexandre, L. P. Corona, D. P. Nunes, J. L. F. Santos, Y. A. d. O. Duarte, and M. L. Lebrão. Gender differences in incidence and determinants of disability in activities of daily living among elderly individuals: Sabe study. *Archives of Gerontology and Geriatrics*, 2012.

M. Allard, J.-P. Cresta, and J.-C. Rochet. Pooling and separating equilibria in insurance markets with adverse selection and distribution costs. *The Geneva Papers on Risk and Insurance Theory*, 22:2:103–120, 1997.

BIBLIOGRAPHY

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders. 3rd ed. Revised.* American Psychiatric Association, Washington, DC, 1987.

K. Andersen, H. Nielsen, A. Lolk, J. Andersen, I. Becker, and P. Kragh-Sorensen. Incidence of very mild to severe dementia and Alzheimer's disease in Denmark, the Odense study. *Neurology*, 52:85–90, 1999.

A. Antoniou, P. Pharoah, G. McMullan, N. Day, B. Ponder, and D. Easton. Evidence for further breast caner susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology*, 21:1–18, 2001.

A. Antoniou, P. Pharoah, G. McMullan, N. Day, M. Stratton, J. Peto, B. Ponder, and D. Easton. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer*, 86:76–83, 2002.

A. Antoniou, P. Pharoah, S. Narod, H. Risch, J. Eyfjord, J. Hopper, N. Loman, H. Olsson, O. Johannsson, A. Borg, et al. Average risks of breast and ovarian cancer associated with brca1 or brca2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *American Journal of Human Genetics*, 72:5:1117–1130, 2003.

A. C. Antoniou, A. B. Spurdle, O. M. Sinilnikova, S. Healey, K. A. Pooley, R. K. Schmutzler, B. Versmold, C. Engel, A. Meindl, N. Arnold, W. Hofmann, C. Sutter, D. Niederacher, H. Deissler, T. Caldes, K. Kampjarvi, H. Nevanlinna, J. Simard, J. Beesley, X. Chen, S. L. Neuhausen, T. R. Rebbeck, T. Wagner, H. T. Lynch, C. Isaacs, J. Weitzel, P. A. Ganz, M. B. Daly, G. Tomlinson, O. I. Olopade, J. L. Bium, F. J. Couch, P. Peterlongo, S. Manoukian, M. Barile, P. Radice, C. I. Szabo, L. H. M. Pereira, M. H. Greene, G. Rennert, F. Leibkowicz, O. Barnett-Griness, I. L. Andrulis, H. Ozcelik, A.-M. Gerdes, M. A. Caligo, Y. Laitman, B. Kaufman, R. Milgrom, E. Friedman, S. M. Domchek, K. L. Nathanson, A. Osorio, G. Llort, R. L. Milne, J. Benitez, U. Hamann, F. B. L. Hogervorst, P. Manders, M. J. L. Ligtenberg, A. M. W. van den Ouweland, S. Peock, M. Cook, R. Platte, D. G. Evans, R. Eeles, G. Pichert, C. Chu, D. Eccles, R. Davidson, F. Douglas, A. K. Godwin, L. Barjhoux, S. Mazoyer, H. Sobol, V. Bourdon, F. Eisinger, A. Chompret, C. Capoulade, B. B.-d. Paillerets, G. M. Lenoir, M. Gauthier-Villars, C. Houdayer, D. Stoppa-Lyonnet, D. F. Easton, K. C. C. Res, OCGN, Swedish BRCA1 and DNA-HEBON Collaborators, EMBRACE, GEMO, and CIMBA. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *American Journal of Human Genetics*, 82:937–948, 2008.

K. Armstrong, B. Weber, G. FitzGerald, J. Hershey, M. Pauly, K. Lemaire, J. Subramanian, and A. D.A. Life insurance and breast cancer risk assessment: adverse

selection, genetic testing decisions, and discrimination. *American Journal of Medical Genetics Part A*, 120A:3:359–364, 2003.

J. E. Bailey-Wilson and A. F. Wilson. Linkage analysis in the next-generation sequencing era. *Human Heredity*, 72:4:228–236, 2011.

C. Ballard, C. O'Brien, R. Barber, D. Neill, and I. McKeith. The progression of cognitive impairment in dementia with Lewy bodies, vascular dementia and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, 16:499–503, 2001.

L. L. Barclay, A. Zemcov, J. P. Blass, and J. Sansone. Survival in alzheimer's disease and vascular dementias. *Neurology*, 35(6):834–834, 1985.

L. Battistin and A. Cagnin. Vascular cognitive disorder. a biological and clinical overview. *Neurochemical Research*, 35:12:1933–1938, 2010.

C. Berr, J. Wancata, and K. Ritchie. Prevalence of dementia in the edlerly in Europe. *European Neuropsychopharmacology*, 15:463–471, 2005.

S. Bochkanov and V. Bystritsky. Alglib, 2013. www.alglib.net.

A. Börsch-Supan, K. Hank, and H. Jürges. A new comprehensive and international view on ageing: introducing the survey of health, ageing and retirement in europe. *European Journal of Ageing*, 2:4:245–253, 2005.

J. R. Botkin, R. T. Croyle, K. R. Smith, B. J. Baty, C. Lerman, D. E. Goldgar, J. M. Ward, B. J. Flick, and J. E. Nash. A model protocol for evaluating the behavioral and psychosocial effects of brca1 testing. *Journal of the National Cancer Institute*, 88:13:872–882, 1996.

L. Bracco, R. Gallato, F. Grigoletto, A. Lippi, V. Lepore, G. Bino, M. P. Lazzaro, F. Carella, T. Piccolo, C. Pozzilli, et al. Factors affecting course and survival in alzheimer's disease: a 9-year longitudinal study. *Archives of Neurology*, 51:12:1213–1019, 1994.

M. Brown. Evidence of adverse selection in the individual health insurance market. *The Journal of Risk and Insurance*, 59:13–33, 1992.

A. Burns. *Ageing and dementia: a methodological approach*, chapter 7: Survival in dementia, pages 125–143. E. Arnold, London, 1993. ISBN 0340568380.

A. Burns, G. Lewis, R. Jacoby, and R. Levy. Factors affecting survival in alzheimers disease. *Psychological Medicine*, 21:2:363–370, 1991.

P. R. Burton, L. J. Palmer, K. Jacobs, K. J. Keen, J. M. Olson, and R. C. Elston. Ascertainment adjustment: where does it take us? *The American Journal of Human Genetics*, 67:6:1505–1514, 2000.

J. K. Cataldo, J. J. Prochaska, and S. A. Glanz. Cigarette smoking is a risk factor for alzheimer's disease: an analysis controlling for tobacco industry affiliation. *Journal of Alzheimers Disease*, 19:2:465–489, 2010.

V. Chandra and R. Pandav. Gene-environment interaction in alzheimer's disease: a potential role for cholesterol. *Neuroepidemiology*, 17:5:225 – 232, 1998.

E. B. Claus, N. Risch, and W. D. Thompson. Autosomal dominant inheritance of early-onset breast cancer. implications for risk prediction. *Cancer*, 73:3:643–651, 1994.

M. P. Coleman, P. Babb, P. Damiecki, P. Grosclaude, S. Honjo, J. Jones, G. Knerer, A. Pitard, M. Quinn, A. Sloggett, and B. De Stavola. *Cancer survival trends in England and Wales, 1971-1995: deprivation and NHS region.* HM Stationery Office, London, 1999.

G. Colgan. Analysis of medical interventions in Alzheimer's disease as the affect long-term care. University College Dublin, 2006.

A. Comas-Herrera, R. Wittenberg, L. Pickard, M. Knapp, and MRC-CFAS. Cognitive impairment in older people: its implications for future demand for services and costs, 2003. PSSRU Discussion Paper 1782/2.

Commission on Funding of Care and Support. Call for evidence on the future funding of care and support, July 2010.

Commission on Funding of Care and Support. Fairer care funding, July 2011.

Continuous Mortality Investigation. Continuous mortality investigation reports number 10, 1990.

Continuous Mortality Investigation. Continuous mortality investigation reports number 23, 2009.

Continuous Mortality Investigation. The CMI Mortality Projections Model, CMI_2011, September 2011. Working Paper 55.

K. Coon, A. Myers, D. Craig, J. Webster, J. Pearson, D. Lince, V. Zismann, T. Beach, D. Leung, L. Bryden, R. Halperin, L. Marlowe, M. Kaleem, D. Walker, R. Ravid, C. Heward, J. Rogers, A. Papassotiropoulos, E. Reiman, J. Hardy, and D. Stephan. A high-density whole-genome association study reveals that APOE is the major

susceptibility gene for sporadic late-onset alzheimer's disease. *Journal of Clinical Psychiatry*, 68:4:613–618, 2007.

J. Copeland, C. McCracken, M. Dewey, K. Wilson, M. Doran, C. Gilmore, A. Scott, and B. Larkin. Undifferentiated dementia, Alzheimer's disease and vascular dementia: age and gender-related incidence in Liverpool - the MRC-ALPHA study. *British Journal of Psychiatry*, 175:433–438, 1999.

M. Corrada, R. Brookmeyer, and C. Kawas. Sources of variability in prevalence rates of alzheimer's disease. *International Journal of Epidemiology*, 24:5:1000–1005, 1995.

C. Courbage and N. Roudaut. Empirical evidence on long-term care insurance purchase in France. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 33:4:645–658, 2008.

A. Cox, A. M. Dunning, M. Garcia-Closas, S. Balasubramanian, M. W. Reed, K. A. Pooley, S. Scollen, C. Baynes, B. A. Ponder, S. Chanock, J. Lissowska, L. Brinton, B. Peplonska, M. C. Southey, J. L. Hopper, M. R. McCredie, G. G. Giles, O. Fletcher, N. Johnson, I. dos Santos Silva, L. Gibson, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, D. Torres, U. Hamann, C. Justenhoven, H. Brauch, J. Chang-Claude, S. Kropp, A. Risch, S. Wang-Gohrke, P. Schuermann, N. Bogdanova, T. Doerk, R. Fagerholm, K. Aaltonen, C. Blomqvist, H. Nevanlinna, S. Seal, A. Renwick, M. R. Stratton, N. Rahman, S. Sangrajrang, D. Hughes, F. Odefrey, P. Brennan, A. B. Spurdle, G. Chenevix-Trench, J. Beesley, A. Mannermaa, J. Hartikainen, V. Kataja, V.-M. Kosma, F. J. Couch, J. E. Olson, E. L. Goode, A. Broeks, M. K. Schmidt, F. B. L. Hogervorst, L. J. Van't Veer, D. Kang, K.-Y. Yoo, D.-Y. Noh, S.-H. Ahn, S. Wedren, P. Hall, Y.-L. Low, J. Liu, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, A. J. Sigurdson, D. L. Stredrick, B. H. Alexander, J. P. Struewing, P. D. Pharoah, D. F. Easton, Kathleen Cunningham Fdn Consortium, and Breast Canc Assoc Consortium. A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*, 39:3: 352–358, 2007.

M. J. Crawley. *The R book*. John Wiley & Sons, 2012.

C. Daykin, D. Akers, A. Macdonald, T. McGleenan, D. Paul, and P. Turvey. Genetics and insurance–some social policy issues. *British Actuarial Journal*, 9:4:787–830, 2003.

R. De Bont and N. van Larebeke. Endogenous dna damage in humans: a review of quantitative data. *Mutagenesis*, 19:3:169–185, 2004.

BIBLIOGRAPHY

L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

H. Diesfeldt, L. Houte, and R. Moerkens. Duration of survival in senile dementia. *Acta Psychiatrica Scandinavica*, 73:4:366–371, 1986.

A. Dinani, D. Grimshaw, N. Robjohns, S. Somerville, A. Spry, and J. Staffurth. A critical review: report of the critical illness healthcare study group, 2000.

N. A. Doherty and P. D. Thistle. Adverse selection with endogenous information in insurance markets. *Journal of Public Economics*, 63:1:83–102, 1996.

D. F. Easton, K. A. Pooley, A. M. Dunning, P. D. P. Pharoah, D. Thompson, D. G. Ballinger, J. P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C. S. Healey, R. Bowman, K. B. Meyer, C. A. Haiman, L. K. Kolonel, B. E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C.-Y. Shen, P.-E. Wu, H.-C. Wang, D. Eccles, D. G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M. R. Stratton, N. Rahman, G. Chenevix-Trench, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K.-Y. Yoo, D.-Y. Noh, S.-H. Ahn, D. J. Hunter, S. E. Hankinson, D. G. Cox, P. Hall, S. Wedren, J. Liu, Y.-L. Low, N. Bogdanova, P. Schuermann, T. Doerk, R. A. E. M. Tollenaar, C. E. Jacobi, P. Devilee, J. G. M. Klijn, A. J. Sigurdson, M. M. Doody, B. H. Alexander, J. Zhang, A. Cox, I. W. Brock, G. MacPherson, M. W. R. Reed, F. J. Couch, E. L. Goode, J. E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J. L. Hopper, M. McCredie, M. Southey, G. G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y.-D. Ko, A. B. Spurdle, J. Beesley, X. Chen, A. Mannermaa, V.-M. Kosma, V. Kataja, J. Hartikainen, N. E. Day, D. R. Cox, B. A. J. Ponder, SEARCH Collaborators, kConFab, and AOCS Management Grp. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447:1087–1095, 2007.

European Court of Justice. C-236/09 Association Belge des Consommateurs Test-Achats ASBL, 2011.

L. A. Farrer, M. F. Brin, A. Goate, J. Kennedy, R. Mayeux, R. H. Myers, P. Reilly, and N. J. Risch. Statement on use of apolipoprotein E testing for Alzheimer disease. *Journal of the American Medical Association*, 274:20:1627–1629, 1995.

L. A. Farrer, L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, R. H. Myers, M. A. Pericak-Vance, N. Risch, and C. M. van Duijn. Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer

disease. a meta-analysis. apoe and alzheimer disease meta analysis consortium. *Journal of the American Medical Association*, 278:16:1349 – 1356, 1997.

A. Finkelstein and K. McGarry. Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economics Review*, 96:938–958, 2006.

A. Finkelstein and J. Poterba. Adverse selection in insurance markets: policyholder evidence from the U.K. annuity market. *Journal of Political Economy*, 112:1:183–208, 2004.

R. Fletcher, editor. *Who cares? The implications of a new partnership to fund long-term care.* The Chartered Insurance Institute, 2011.

M. F. Folstein, S. E. Folstein, and P. R. McHugh. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12:3:189–198, 1975.

D. Ford, D. Easton, M. Stratton, S. Narod, D. Goldgar, P. Devilee, D. Bishop, B. Weber, G. Lenoir, J. Chang-Claude, H. Sobol, M. Teare, J. Struewing, A. Arason, S. Scherneck, J. Peto, T. Rebbeck, P. Tonin, S. Neuhausen, R. Barkardottir, J. Eyfjord, H. Lynch, B. Ponder, S. Gayther, J. Birch, A. Lindblom, D. Stoppa-Lyonnet, Y. Bignon, A. Borg, U. Hamann, N. Haites, R. Scott, C. Maugard, H. Vasen, S. Seitz, L. Cannon-Albright, A. Schofield, and M. Zelada-Hedman. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics*, 62:3:676–689, 1998.

J. Forder and J.-L. Fernández. Analysing the costs and benefits of social care funding arrangements in england: technical report, 2009. PSSRU Discussion Paper 2644.

T. M. R. C. C. Function and A. Study. Cognitive function and dementia in six areas of England and Wales: the distribution of MMSE and prevalence of GMS organicity level in the MRC CFA Study. *Psychological Medicine*, 28:2:319–335, 1998.

M. Gatz, C. Reynolds, L. Fratiglioni, B. Johansson, J. Mortimer, S. Berg, A. Fiske, and N. Pedersen. Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, 63:2:168–74, 2006.

S. Gauthier, B. Vellas, M. Farlow, and D. Burn. Aggressive course of disease in dementia. *Alzheimer's & Dementia*, 2:3:210–217, 2006.

A. Gerrish, G. Russo, A. Richards, V. Moskvina, D. Ivanov, D. Harold, R. Sims, R. Abraham, P. Hollingworth, J. Chapman, M. Hamshere, J. Pahwa, K. Dowzell,

A. Williams, N. Jones, C. Thomas, A. Stretton, A. Morgan, S. Lovestone, J. Powell, P. Proitsi, M. Lupton, C. Brayne, D. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K. Brown, P. Passmore, D. Craig, B. McGuinness, S. Todd, J. Johnston, C. Holmes, D. Mann, A. Smith, S. Love, P. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, H. Kölsch, R. Heun, B. Schürmann, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, M. Hüll, D. Rujescu, A. Goate, J. Kauwe, C. Cruchaga, P. Nowotny, J. Morris, K. Mayo, G. Livingston, N. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, G. Davies, S. Harris, J. Starr, I. Deary, A. Al-Chalabi, C. Shaw, M. Tsolaki, A. Singleton, R. Guerreiro, T. Mühleisen, M. Nöthen, S. Moebus, K. Jöckel, N. Klopp, H. Wichmann, M. Carrasquillo, V. Pankratz, S. Younkin, L. Jones, P. Holmans, M. O'Donovan, M. Owen, and J. Williams. The role of variation at AβPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. *Journal of Alzheimer's Disease*, 28:2:377–387, 2012.

M. Ghoussaini, O. Fletcher, K. Michailidou, C. Turnbull, M. K. Schmidt, E. Dicks, J. Dennis, Q. Wang, M. K. Humphreys, C. Luccarini, C. Baynes, D. Conroy, M. Maranian, S. Ahmed, K. Driver, N. Johnson, N. Orr, I. dos Santos Silva, Q. Waisfisz, H. Meijers-Heijboer, A. G. Uitterlinden, F. Rivadeneira, P. Hall, K. Czene, A. Irwanto, J. Liu, H. Nevanlinna, K. Aittomaki, C. Blomqvist, A. Meindl, R. K. Schmutzler, B. Muller-Myhsok, P. Lichtner, J. Chang-Claude, R. Hein, S. Nickels, D. Flesch-Janys, H. Tsimiklis, E. Makalic, D. Schmidt, M. Bui, J. L. Hopper, C. Apicella, D. J. Park, M. Southey, D. J. Hunter, S. J. Chanock, A. Broeks, S. Verhoef, F. B. Hogervorst, P. A. Fasching, M. P. Lux, M. W. Beckmann, A. B. Ekici, E. Sawyer, I. Tomlinson, M. Kerin, F. Marme, A. Schneeweiss, C. Sohn, B. Burwinkel, P. Guenel, T. Truong, E. Cordina-Duverger, F. Menegaux, S. E. Bojesen, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, R. L. Milne, M. R. Alonso, A. Gonzalez-Neira, J. Benitez, H. Anton-Culver, A. Ziogas, L. Bernstein, C. C. Dur, H. Brenner, H. Muller, V. Arndt, C. Stegmaier, C. Justenhoven, H. Brauch, T. Bruning, S. Wang-Gohrke, U. Eilber, T. Dork, P. Schurmann, M. Bremer, P. Hillemanns, N. V. Bogdanova, N. N. Antonenkova, Y. I. Rogov, J. H. Karstens, M. Bermisheva, D. Prokofieva, E. Khusnutdinova, A. Lindblom, S. Margolin, A. Mannermaa, V. Kataja, V.-M. Kosma, J. M. Hartikainen, D. Lambrechts, B. T. Yesilyurt, G. Floris, K. Leunen, S. Manoukian, B. Bonanni, S. Fortuzzi, P. Peterlongo, F. J. Couch, X. Wang, K. Stevens, A. Lee, G. G. Giles, L. Baglietto, G. Severi, C. McLean, G. G. Alnaes, V. Kristensen, A.-L. Borrensen-Dale, E. M. John, A. Miron, R. Winqvist, K. Pylkas, A. Jukkola-Vuorinen, S. Kauppila, I. L. Andrulis, G. Glendon, A. M. Mulligan, P. Devilee, C. J. van Asperen, R. A. Tollenaar, C. Seynaeve, J. D. Figueroa, M. Garcia-Closas, L. Brinton, J. Lissowska, M. J. Hooning, A. Hollestelle, R. A. Oldenburg, A. M. van den Ouweland,

A. Cox, M. W. Reed, M. Shah, A. Jakubowska, J. Lubinski, K. Jaworska, K. Durda, M. Jones, M. Schoemaker, A. Ashworth, A. Swerdlow, J. Beesley, X. Chen, K. R. Muir, A. Lophatananon, S. Rattanamongkongul, A. Chaiwerawattana, D. Kang, K.-Y. Yoo, D.-Y. Noh, C.-Y. Shen, J.-C. Yu, P.-E. Wu, C.-N. Hsiung, A. Perkins, R. Swann, L. Velentzis, D. M. Eccles, W. J. Tapper, S. M. Gerty, N. J. Graham, B. A. Ponder, G. Chenevix-Trench, P. D. Pharoah, M. Lathrop, A. M. Dunning, N. Rahman, J. Peto, and D. Easton. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nature Genetics*, 44:3:312–318, 2012.

D. E. Goldgar, P. Fields, C. M. Lewis, T. D. Tran, L. A. Cannon-Albright, J. H. Ward, J. Swensen, and M. H. Skolnick. A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis. *Journal of the National Cancer Institute*, 86:3:200–209, 1994.

P. Graham, K. Mengersen, and A. Morton. Confidence limits for the ratio of two rates based on likelihood scores: Non-iterative method. *Statistics in Medicine*, 22: 2071–2083, 2003.

E. H. Gui and A. Macdonald. Early-onset alzheimer's disease, critical illness and life insurance. *Genetics and Insurance Centre Research Report*, 02:2:1–31, 2002.

E. H. Gui, B. Lu, A. S. Macdonald, H. R. Waters, and C. T. Wekwete. The genetics of breast and ovarian cancer III: a new model of family history with insurance applications. *Scandinavian Actuarial Journal*, 6:413–419, 2006.

M. Guillén and A. Comas-Herrera. How much risk is mitigated by LTC protection schemes? a methodological note and a case study of the public system in Spain. *The Geneva Papers on Risk and Insurance—Issues and Practice*, 37:4:712–724, 2012.

C. Gutiérrez and A. S. Macdonald. Huntington's disease and insurance i: A model of huntington's disease. *Genetics and Insurance Centre Research Report*, 02/3:1–28, 2002.

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

C. Gutiérrez and A. S. Macdonald. Huntington's disease, critical illness and life insurance. *Scandinavian Actuarial Journal*, 4:279–313, 2004.

C. Gutiérrez and A. S. Macdonald. Adult polycystic kidney disease and insurance: a case study in genetic heterogeneity. *North American Actuarial Journal*, 11:1: 90–118, 2007.

R. Hancock, R. Wittenberg, L. Pickard, A. Comas-Herrera, A. Juarez-Garcia, D. King, and J. Malley. Paying for long-term care for older people in the UK: modelling the costs and distributional effects of a range of options, 2007. PSSRU Discussion Paper 2336/2.

P. Hartge, A. S. Whittemore, J. Itnyre, L. McGowan, D. Cramer, and The Collaborative Ovarian Cancer Group. Rates and risks of ovarian cancer in subgroups of white women in the united states. *Obstetrics & Gynecology*, 84:5:760–764, 1994.

A. Heyman, B. Peterson, G. Fillenbaum, and C. Pieper. The consortium to establish a registry for alzheimer's disease (cerad). part xiv demographic and clinical predictors of survival in patients with alzheimer's disease. *Neurology*, 46:3:656–660, 1996.

A. Hofman, D. Grobbee, P. De Jong, and F. Van den Ouweland. Determinants of disease and disability in the elderly: the rotterdam elderly study. *European Journal of Epidemiology*, 7:4:403–422, 1991.

A. Howell, S. Astley, J. Warwick, P. Stavrinos, S. Sahin, S. Ingham, H. McBurney, B. Eckersley, M. Harvie, M. Wilson, U. Beetles, R. Warren, A. Hufton, J. Sergeant, W. Newman, I. Buchan, J. Cuzick, and D. G. Evans. Prevention of breast cancer in the context of a national breast screening programme. *Journal of Internal Medicine*, 271:4:321–330, 2012.

M. Hoy and M. Polborn. The value of genetic information in the life insurance market. *Journal of Public Economics*, 78:3:235–252, 2000.

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

Y. Huang and L. Mucke. Alzheimer mechanisms and therapeutic strategies. *Cell*, 148:6:1204–1222, 2012.

D. J. Hunter, P. Kraft, K. B. Jacobs, D. G. Cox, M. Yeager, S. E. Hankinson, S. Wacholder, Z. Wang, R. Welch, A. Hutchinson, J. Wang, K. Yu, N. Chatterjee, N. Orr, W. C. Willett, G. A. Colditz, R. G. Ziegler, C. D. Berg, S. S. Buys, C. A. McCarty, H. S. Feigelson, E. E. Calle, M. J. Thun, R. B. Hayes, M. Tucker, D. S. Gerhard, J. F. Fraumeni, Jr., R. N. Hoover, G. Thomas, and S. J. Chanock. A genome-wide association study identifies alleles in fgfr2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics*, 39:870–874, 2007.

H. Iwasa, Y. Gondo, Y. Yoshida, J. Kwon, H. Inagaki, C. Kawaai, Y. Masui, H. Kim, H. Yoshida, and T. Suzuki. Cognitive performance as a predictor of functional

decline among the non-disabled elderly dwelling in a japanese community: A 4-year population-based prospective cohort study. *Archives of Gerontology and Geriatrics*, 47:1:139–149, 2008.

K. Jones. The cost of providing home care. In *Unit costs of health & social care 2005*. University of Kent, Personal Social Services Research Unit, Kent, 2006. ISBN 1-902-671-40-6.

A. Jorm and D. Jolley. The incidence of dementia: a meta-analysis. *Neurology*, 51:3: 728–733, 1998.

D. Kahneman and A. Tversky. Prospect theory: an analysis of decision under risk. *Econometrica*, 47:2:263–291, 1979.

S. Katz, T. D. Downs, H. R. Cash, and R. C. Grotz. Progress in development of the index of adl. *The Gerontologist*, 10:1:20–30, 1970.

B. M. Kofler, E. A. Miles, P. Curtis, C. K. Armah, S. Tricon, J. Grew, F. L. Napper, L. Farrell, G. Lietz, C. J. Packard, M. J. Caslake, J. C. Mathers, C. M. Williams, P. C. Calder, and A. M. Minihane. Apolipoprotein E genotype and the cardiovascular disease risk phenotype: Impact of sex and adiposity (the FINGEN study). *Atherosclerosis*, 221:2:467–470, 2012.

K. Lange. An approximate model of polygenic inheritance. *Genetics*, 147:1423–1430, 1997.

L. Launer, K. Andersen, M. Dewey, L. Letenneur, A. Ott, L. Amaducci, C. Brayne, J. Copeland, J. Dartigues, P. Kragh-Sorensen, A. Lobo, J. Martinez-Lage, T. Stijnen, A. Hofman, and EURODEM Incidence Res Grp Work Grp. Rates and risk factors for dementia and Alzheimer's disease results from EURODEM pooled analyses. *Neurology*, 52:1:78–84, 1999.

L. J. Launer. Overview of incidence studies of dementia conducted in Europe. *Neuroepidemiology*, 11 Suppl 1:2–13, 1992.

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

L. Letenneur, D. Commenges, J. Dartigues, and P. Barberger-Gateau. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western france. *International Journal of Epidemiology*, 23:6:1256–1261, 1994.

J. Lloyd. Gone for good? pre-funded insurance for long-term care. Strategic Society Centre, 2011.

G. Loewenstein. Emotions in economic theory and economic behavior. *The American Economic Review*, 90:2:426–432, 2000.

L.-F. Low, M. H. Yap, and H. Brodaty. Will testing for apolipoprotein E assist in tailoring dementia risk reduction? a review. *Neuroscience and Biobehavioral Reviews*, 34:3:408–437, 2010.

B. Lu, A. S. Macdonald, and H. R. Waters. The genetics of breast and ovarian cancer IV: A model of breast cancer progression. *Scandinavian Actuarial Journal*, 6:239–266, 2011a.

B. Lu, A. S. Macdonald, H. R. Waters, and F. Yu. The genetics of breast and ovarian cancer V: Application to income protection insurance. *Scandinavian Actuarial Journal*, 4:267–291, 2011b.

L. Lu, A. S. Macdonald, H. R. Waters, and F. Yu. A dynamic familyhistory model of hereditary nonpolyposis colorectal cancer and critical illness insurance. *Annals of Actuarial Science*, 2:2:289–325, 2007.

L. MacCalman. *Effects of genetic testing on insurance:Pedigree analysis and ascertainment adjustment*. PhD thesis, Heriot-Watt University, 2009.

A. Macdonald and F. Yu. The impact of genetic information on the insurance industry: conclusions from the bottom-up modelling programme. *ASTIN Bulletin*, 41:343–376, 2011.

A. S. Macdonald. How will improved forecasts of individual lifetimes affect underwriting. *Philosophical Transactions of the Royal Society B*, 352:1067–1075, 1997.

A. S. Macdonald and K. McIvor. Application of a polygenic model of breast and ovarian cancer to critical illness insurance. *Annals of Actuarial Science*, 1:2:319–343, 2006.

A. S. Macdonald and K. McIvor. Modelling adverse selection in the presence of a common genetic disorder: The breast cancer polygene. *ASTIN Bulletin*, 39:373–402, 2009.

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

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

A. S. Macdonald and P. Tapadar. Multifactorial genetic disorders and adverse selection: epidemiology meets economics. *Journal of Risk and Insurance*, 77:1:155–182, 2010.

A. S. Macdonald, H. R. Waters, and C. T. Wekwete. The genetics of breast and ovarian cancer I: a model of family history. *Scandinavian Actuarial Journal*, 2003: 1–27, 2003a.

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

F. I. Mahoney and D. W. Barthel. Functional evaluation: the barthel index. *Maryland State Medical Journal*, 14:61–65, 1965.

T. A. Manolio, R. L. Chisholm, B. Ozenberger, D. M. Roden, M. S. Williams, R. Wilson, D. Bick, E. P. Bottinger, M. H. Brilliant, C. Eng, et al. Implementing genomic medicine in the clinic: the future is here. *Genetics in Medicine*, 15:4:258–267, 2013.

K. G. Manton. A longitudinal study of functional change and mortality in the united states. *Journal of Gerontology*, 43:5:S153–S161, 1988.

F. E. Matthews, A. Arthur, L. E. Barnes, J. Bond, C. Jagger, L. Robinson, and C. Brayne. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of england: results of the cognitive function and ageing study i and {II}. *The Lancet*, Online:1–8, 2013.

A. McCormick, D. Fleming, and J. Charlton. Morbidity statistics from general practice: Fourth national study 1991–1992, 1995.

L. C. McGuire, E. S. Ford, and U. A. Ajani. Cognitive functioning as a predictor of functional disability in later life. *American Journal of Geriatric Psych*, 14:1:36–42, 2006.

P. McNamee. A comparison of the grade of membership measure with alternative health indicators in explaining costs for older people. *Health Economics*, 13:379–395, 2004.

G. A. Melnick and K. Fonkych. Hospital pricing and the uninsured: Do the uninsured pay higher prices? *Health Affairs*, 27:2:w116–w122, 2008.

W. Mendenhall. *Mathematical statistics with applications*. Duxbury Press, Boston, 1986. ISBN 0-87150-939-3.

S. Narod, S. Neuhausen, G. Vichodez, S. Armel, H. Lynch, P. Ghadirian, S. Cummings, O. Olopade, D. Stoppa-Lyonnet, F. Couch, T. Wagner, E. Warner, W. Foulkes, H. Saal, J. Weitzel, A. Tulman, A. Poll, R. Nam, P. Sun, and Hereditary Breast Cancer Study Grp. Rapid progression of prostate cancer in men with a BRCA2 mutation. *British Journal of Cancer*, 99:2:371–374, 2008.

R. Neale, C. Brayne, and A. Johnson. Cognition and survival: an exploration in a large multicentre study of the poulation aged 65 years and over. *International Journal of Epidemiology*, 30:6:1383–1388, 2001.

R. Norberg. Differential equations for moments of present values in life insurance. *Insurance: Mathematics and Economics*, 17:2:171–180, 1995.

S. R. Nuttall, R. J. L. Blackwood, B. M. H. Bussell, J. P. Cliff, M. J. Cornall, A. Cowley, P. L. Gatenby, and J. M. Webber. Financing long-term care in great britain. *Journal of the Institute of Actuaries*, 121:1:1–68, 1994. ISSN 00202681.

Office for National Statistics. Cancer 1971–1997. CD-ROM, 1999.

Office for National Statistics. 2010-based national population projections. `http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-229071`, 2011.

Office for National Statistics. Population estimates for uk, england and wales, scotland and northern ireland, population density tables, 1981 to 2010. `http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-268725`, 2012a.

Office for National Statistics. Mortality in england and wales, 2012b.

Office for National Statistics. Average weekly earnings dataset. `http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/january-2013/dataset--earnings.html`, 2013.

E. Oster, I. Shoulson, K. Quaid, and E. R. Dorsey. Genetic adverse selection: Evidence from long-term care insurance and Huntington disease. *Journal of Public Economics*, 94:1041–1050, 2010.

A. Ott, M. M. Breteler, F. van Harskamp, J. J. Claus, T. J. D. E. G. van der Cammen, and A. Hofman. Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. *British Medical Journal*, 310: 970–973, 1995.

D. Parkin, F. Bray, and S. Devesa. Cancer burden in the year 2000. the global picture. *European Journal of Cancer*, 37, Supplement 8:4–66, 2001.

G. Parmigiani, D. A. Berry, and O. Aguilar. Determining carrier probabilities for breast cancer–susceptibility genes brca1 and brca2. *The American Journal of Human Genetics*, 62:1:145–158, 1998.

J. J. Pasternak. *An introduction to human molecular genetics: mechanisms of inherited diseases.* Fitzgerald, 1999. ISBN 1891786032.

M. V. Pauly. The rational nonpurchase of long-term-care insurance. *Journal of Political Economy*, pages 153–168, 1990.

P. D. Pharoah, A. C. Antoniou, D. F. Easton, and B. A. Ponder. Polygenes, risk prediction, and targeted prevention of breast cancer. *New England Journal of Medicine*, 358:2796–2803, 2008.

D. Pritchard. *The genetics of Alzheimer's disease and adverse selection in the long-term care insurance market.* PhD thesis, Heriot-Watt University, 2002.

D. Pritchard. Modeling disability in long-term care insurance. *North American Actuarial Journal*, 10:4:48–75, 2006.

B. Rickayzen and D. Walsh. A multi-state model of disability for the United Kingdom: Implications for the future need for long term care for the elderly. *British Actuarial Journal*, 8:11:341–393, 2002.

M. Ropka, J. Wenzel, E. Phillips, M. Siadaty, and J. Philbrick. Uptake rates for breast cancer genetic testing: A systematic review. *Cancer Epidemiology, Biomarkers, and Prevention*, 15:5:840–855, 2006.

M. Rothschild and J. Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, pages 629–649, 1976.

Royal Commission on Long Term Care. *With respect to old age: long term care: rights and responsibilities*, volume 1. Stationery Office/Tso, 1999.

K. S. Santacruz and D. Swagerty. Early diagnosis of dementia. *American Family Physician*, 63:4:703–713, 2001.

A. Saunders, W. Strittmatter, D. Schmechel, P. S. George-Hyslop, M. Pericak-Vance, S. Joo, B. Rosi, J. Gusella, D. Crapper-MacLachlan, M. Alberts, Hulette, B. Crain, D. Goldgaber, and A. Roses. Association of apolipoprotein e allele 4 with late-onset familial and sporadic alzheimer's disease. *Neurology*, 43:8:1467–1467, 1993.

R. B. Sayetta. Rates of senile dementia -- alzheimer's type in the baltimore longitudinal study. *Journal of Chronic Diseases*, 39:4:271–286, 1986.

C. Shaw. Fertility assumptions for 1989-based population projections for england and wales. *Population Trends*, 61:17–23, 1990.

P. Siegelman. Adverse selection in insurance markets: an exaggerated threat. *Yale Law Journal*, 113:1223–1281, 2003.

A.-B. Skytte, A.-M. Gerdes, M. Andersen, L. Sunde, K. Brøndum-Nielsen, M. Waldstrøm, S. Kølvraa, and D. Crüger. Risk-reducing mastectomy and salpingo-oophorectomy in unaffected brca mutation carriers: Uptake and timing*. *Clinical Genetics*, 77:4:342–349, 2010.

Y. Song, M. J. Stampfer, and S. Liu. Meta-analysis: apolipoprotein e genotypes and risk for coronary heart disease. *Annals of Internal Medicine*, 141:2:137–147, 2004.

A. Spiers, R. Matthews, C. Jagger, F. Matthews, C. Boult, and T. Robinson. Diseases and impairment as risk factors for onset of disability in the older population in England and Wales: Findings from the MRC Cognitive Function and Ageing Study (MRC CFAS). *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 60:2:248–254, 2005.

S. N. Stacey, A. Manolescu, P. Sulem, T. Rafnar, J. Gudmundsson, S. A. Gudjonsson, G. Masson, M. Jakobsdottir, S. Thorlacius, A. Helgason, K. K. Aben, L. J. Strobbe, M. T. Albers-Akkers, D. W. Swinkels, B. E. Henderson, L. N. Kolonel, L. Le Marchand, E. Millastre, R. Andres, J. Godino, M. D. Garcia-Prats, E. Polo, A. Tres, M. Mouy, J. Saemundsdottir, V. M. Backman, L. Gudmundsson, K. Kristjansson, J. T. Bergthorsson, J. Kostic, M. L. Frigge, F. Geller, D. Gudbjartsson, H. Sigurdsson, T. Jonsdottir, J. Hrafnkelsson, J. Johannsson, T. Sveinsson, G. Myrdal, H. N. Grimsson, T. Jonsson, S. von Holst, B. Werelius, S. Margolin, A. Lindblom, J. I. Mayordomo, C. A. Haiman, L. A. Kiemeney, O. T. Johannsson, J. R. Gulcher, U. Thorsteinsdottir, A. Kong, and K. Stefansson. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*, 39:7:865–869, 2007.

J. A. Stewart, R. Dundas, R. Howard, A. Rudd, and C. Wolfe. Ethnic differences in incidence of stroke: prospective study with stroke register. *British Medical Journal*, 318:7189:967–971, 1999.

K. Subramanian, J. Lemaire, J. Hershey, M. Pauly, K. Armstrong, and D. Asch. Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *Journal of Risk and Insurance*, 66:4:531–550, 1999.

P. Sudbery. *Human Molecular Genetics (Second Edition)*. Prentice Hall, Harlow, England, 2002. ISBN 0-13-042811-6.

Y. C. Tai, S. Domchek, G. Parmigiani, and S. Chen. Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *Journal of the National Cancer Institute*, 99:23:1811–1814, 2007.

Ü. Taş, A. P. Verhagen, S. Bierma-Zeinstra, A. Hofman, E. Odding, H. A. Pols, and B. W. Koes. Incidence and risk factors of disability in the elderly: the rotterdam study. *Preventive medicine*, 44(3):272–278, 2007.

D. H. Taylor, R. M. Cook-Deegan, S. Hiraki, J. S. Roberts, D. G. Blazer, and R. C. Green. Genetic testing for alzheimers and long-term care insurance. *Health Affairs*, 29:1:102–108, 2010.

A. Tutt and A. Ashworth. The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. *Trends in Molecular Medicine*, 8:12:571–579, 2002.

A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992.

K. S. Viswanathan, J. Lemaire, K. Withers, K. Armstrong, A. Baumritter, J. C. Hershey, M. V. Pauly, and D. A. Asch. Adverse selection in term life insurance purchasing due to the brca1/2 genetic test and elastic demand. *Journal of Risk and Insurance*, 74:1:65–86, 2007.

J. S. Walsh, H. G. Welch, and E. B. Larson. Survival of outpatients with alzheimer-type dementia. *Annals of Internal Medicine*, 113:6:429–434, 1990.

H. Ward, P. N. Mitrou, R. Bowman, R. Luben, N. J. Wareham, K. Tee Khaw, and S. Bingham. Apoe genotype, lipids, and coronary heart disease risk: A prospective population study. *Archives of Internal Medicine*, 169:15:1424–1429, 2009.

E. Washbrook. Risk factors and epidemiology of breast cancer. *Womens Health Medicine*, 3:1:8–14, 2006.

J. D. Watson and F. H. Crick. Molecular structure of nucleic acids. *Nature*, 171:4356: 737–738, 1953.

R. Willets, A. Gallop, P. Leandro, J. Lu, A. Macdonald, K. Miller, S. Richards, N. Robjohns, J. Ryan, and H. Waters. Longevity in the 21st century. *British Actuarial Journal*, 10:4:685–832, 2004.

R. Wittenberg, L. Pickard, A. Comas-Herrera, B. Davies, and R. Darton. *Demand for long-term care: projections of long-term care finance for elderly people*. Personal Social Services Research Unit, Canterbury, UK, 1998.

R. Wittenberg, A. Comas-Herrera, D. King, J. Malley, L. Pickard, and R. Darton. Future demand for long-term care, 2002 to 2041: projections of demand for long-term care for older people in england, 2006. PSSRU Discussion Paper 2330.

T. Zhou-Richter, M. J. Browne, and H. Gründl. Don't they care? or, are they tust unaware? risk perception and the demand for long-term care insurance. *Journal of Risk and Insurance*, 77:4:715–747, 2010.

C. D. Zick, K. R. Smith, R. N. Mayer, and J. R. Botkin. Genetic testing, adverse selection, and the demand for life insurance. *American Journal of Medical Genetics*, 93:1:29–39, 2000.

C. D. Zick, C. J. Mathews, J. S. Roberts, R. Cook-Deegan, R. J. Pokorski, and R. C. Green. Genetic testing for alzheimers disease and its impact on insurance purchasing behavior. *Health Affairs*, 24:2:483–490, 2005.

Appendix A

Numerical Methods

A.1 Kolmogorov Forward Equations

Throughout this thesis we will use Markov multiple state models to represent the health of our modelled populations. The transition intensity from state i to state j at age $x + t$ is denoted μ_{x+t}^{ij}. Denote the probability a life is in state j at age $x + t$ conditional on being state i at age x is denoted ${}_t p_x^{ij}$. The occupancy probabilities in a Markov model can be found by solving differential equations known as Kolmogorov forward equations:

$$\frac{d}{dt} {}_t p_x^{ij} = \sum_{k \neq j} \left({}_t p_x^{ik} \, \mu_{x+t}^{kj} - {}_t p_x^{ij} \, \mu_{x+t}^{jk} \right), \tag{A.1}$$

with boundary conditions ${}_0 p_x^{ij} = \delta^{ij}$, where δ_{ij} is the Kronecker delta.

In a two-state model with no backward transitions, this reduces to solving

$$\frac{d}{dt} {}_t p_x^{ii} = {}_t p_x^{ii} \, \mu_{x+t}^{ij}, \tag{A.2}$$

so

$$ {}_t p_x^{ii} = 1 - \exp \left(\int_0^t \mu_{x+s}^{ij} \, ds \right). \tag{A.3}$$

However, in a more complicated model, there may be no tractable form and the solution will need to be found numerically.

A.2 Runge-Kutta Method

The simplest method for solving differential equations is Euler's method which approximates the function based on the derivative at the beginning of each step. However, the stability and error for a given step size can be improved by using a method that incorporates an approximate calculation of the derivative at the midpoint of the step. One such method is the 4th-order Runge-Kutta method.

To find a function $y(t)$, with derivative of the form $y' = f(t, y)$ and whose initial value $y(t_0) = y_0$ with the 4th-order Runga-Kutta method and a step size of h, we use iterations

$$y_{n+1} \;=\; y_n + h\tfrac{1}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right)$$

$$t_{n+1} \;=\; t_n + h$$

for n = 0, 1, 2, 3, . . . , using

$$k_1 \;=\; f(t_n, y_n),$$

$$k_2 \;=\; f(t_n + \tfrac{1}{2}h, y_n + h\tfrac{1}{2}k_1),$$

$$k_3 \;=\; f(t_n + \tfrac{1}{2}h, y_n + h\tfrac{1}{2}k_2),$$

$$k_4 \;=\; f(t_n + h, y_n + hk_3).$$

This is repeated over the entire period.

Using a negative value for h, and a terminal condition instead of an initial condition, we can solve differential equations backwards.

A.3 Simpson's Rule

We will face functions which need to be integrated but for which the anti-derivative cannot be written in elementary form. Numerical methods exist to integrate such functions approximately.

By rewriting an integral in the form of an initial value problem $I'(x) = f(x)$ with $I(a) = 0$, and solving with the 4th-order Runge Kutta method, the method reduces to the widely used Simpson's rule:

$$\int_a^b f(x)\,dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]. \tag{A.4}$$

Appendix B

Transition Intensities for Other CI

The model of critical illnesses in Chapter 2 has transitions from healthy to BC, OC, Other CI and Dead. The transition intensities we use for the state Other CI are based on the onset rates from Gutiérrez and Macdonald (2003), with an adjustment to remove BC and OC using the onset rates of Macdonald et al. (2003a).

Gutiérrez and Macdonald (2003) fitted models for cancer, heart attack and stroke for each sex. Heart attacks and strokes often result in death. Stand alone critical illness insurance commonly does not pay out if there is a death within 28 days of illness, so they make adjustment for survival after stroke and heart attack (death from cancer usually occurs more than 28 days after diagnosis). We list their functions here. Let μ_x^c, μ_x^b, μ_x^o, μ_x^h and μ_x^s, denote the onset rates at age x, of cancer, breast cancer, ovarian cancer, heart attack and stroke respectively. Further, denote the probability of survival for 28 days after heart attack and stroke at age x, by p_x^h and p_x^s, respectively.

Cancer

Onset rates for cancer were estimated using cancer registrations in England and Wales between 1990 and 1992 from Office for National Statistics (1999).

For males their fitted model was,

$$\mu_x^c = \begin{cases} \exp\left(-11.25 + 0.105x\right) & \text{if } x < 51 \\ 0.2591585 - 0.0124735x + 0.0001916916x^2 - 8.952933 \times 10^{-7}x^3 & \text{if } x \geq 60. \end{cases}$$

(B.1)

Between the ages 51 and 60 they blend these functions linearly,

$$
\begin{aligned}
\mu^c_{x+t} &= \frac{60-x}{9} \exp\left(-11.25 + 0.105x\right) \\[2mm]
&+ \frac{x-51}{9}\left(0.2591585 - 0.0124735x + 0.0001916916x^2 - 8.952933 \times 10^{-7}x^3\right).
\end{aligned}
\tag{B.2}
$$

The fitted model for females is discontinuous due to the impact of the breast cancer screening programme available to all women over the age of 50:

$$
\mu^c_x = \begin{cases}
\exp\left(-10.78 + 0.123x - 0.00033x^2\right) & \text{if } x < 53 \\[2mm]
-0.01545632 + 0.0003805097x & \text{if } x \geq 53
\end{cases}
\tag{B.3}
$$

Female breast cancer and ovarian cancer onset rates were fitted to the same cancer registration data source as above, by Macdonald et al. (2003a). In the case of breast cancer, this also had a discontinuity at age 53:

$$
\mu'^{BC}_x = \begin{cases}
\dfrac{1}{\Gamma(8.7305)} 0.0742^{8.7305} \exp\left(-0.00742x\right) x^{7.7305} & \text{if } x < 53 \\[3mm]
\begin{aligned}
&0.00012 + 0.00018\left(x-35\right) \\
&\quad - 0.000005\left(x-35\right)^2 + 5.29 \times 10^{-8}\left(x-35\right)^3
\end{aligned} & \text{if } x \geq 53
\end{cases}
\tag{B.4}
$$

For ovarian cancer, it was,

$$
\mu'^{OC}_x = \begin{cases}
\frac{1}{\Gamma(6.92)} 0.035^{6.92} \exp\left(-0.035x\right) x^{5.92} & \text{if } x < 45 \\[2mm]
0.0001554 + 0.000029\left(x-45\right) - 0.00000048\left(x-45\right)^2 & \text{if } x \geq 55.
\end{cases}
\tag{B.5}
$$

Between ages 45 and 55, these were blended in a similar fashion as above,

$$
\begin{aligned}
\mu'^{OC}_x &= \frac{1}{\Gamma(6.92)} 0.035^{6.92} \exp\left(-0.035x\right) x^{5.92} \frac{55-x}{10} \\[2mm]
&+ \left[0.0001554 + 0.000029\left(x-45\right) - 0.00000048\left(x-45\right)^2\right] \frac{x-45}{10}.
\end{aligned}
\tag{B.6}
$$

Heart Attack

Heart onset rates were estimated using the cases of first-ever heart attack from McCormick et al. (1995), data from U.K. general practices in the period 1991–1992. For

Table B.1: 28-day survival probabilities for males, following a heart attack. Source: Dinani et al. (2000), via Gutiérrez and Macdonald (2003).

Age (x)	p_x^h	Age (x)	p_x^h
40–42	0.84	60–61	0.78
43–46	0.83	62–64	0.77
47–52	0.82	65–74	0.76
53–56	0.81	75–79	0.75
57	0.80	80+	0.74

males the fitted rates were,

$$\mu_x^h = \begin{cases} \exp\left(-13.2238 + 0.152568\right) & \text{if } x < 44 \\ -0.01245106 + 0.000315605x & \text{if } x > 49, \end{cases} \tag{B.7}$$

with blending between ages 44 and 49:

$$\begin{aligned} \mu_x^h &= \exp\left(-13.2238 + 0.152568\right) \frac{49 - x}{49 - 44} \\ &+ \left(-0.01245106 + 0.000315605x\right) \frac{x - 44}{49 - 44}. \end{aligned} \tag{B.8}$$

The onset rates for females were fitted as,

$$\mu_x^h = \frac{0.598694}{\Gamma\left(15.6412\right)} 0.15317^{15.6412} \exp\left(-0.15317x\right) x^{14.6412}. \tag{B.9}$$

The 28-day survival probabilities were taken from Dinani et al. (2000). Males follow the probabilities given in Table B.1 and females $p_x^h = 0.79$ for all ages x.

Stroke

The onset rates of stroke were fitted to the ungraduated onset rates of Stewart et al. (1999), a prospective community stroke register of population of 234,533 lives in south London over 1995–1996. Male onset rates were fitted as,

$$\mu_x^s = \exp\left(-16.9524 + 0.294973x - 0.001904x^2 + 0.00000159449x^3\right). \tag{B.10}$$

Female onset rates were fitted as,

$$\mu_x^s = \exp\left(-11.1477 + 0.081076x\right) \tag{B.11}$$

The 28-day survival probability after a stroke, taken from Dinani et al. (2000),

were calculated as

$$p_x^s = \frac{0.9 - 0.002x}{0.9},$$

(B.12)

for males and females.

Total Transition Intensity for Other CI

Gutiérrez and Macdonald (2003) assumed that critical illness claims due to other diseases amounted to 15% of the claims due to cancer, heart attack and stroke. The transition intensity from Healthy to Other CI in our model (depicted in Figure 2.1) can be calculated as,

$$\mu^{CI}(x+t) = 1.15 \left(\mu_{x+t}^c + p_{x+t}^h \mu_x^h + p_{x+t}^s \mu_{x+t}^s \right) - \mu_{x+t}^b - \mu_{x+t}^o.$$

(B.13)

Appendix C

CMI Mortality Improvements Model 2011 Assumptions

In this appendix we list the assumptions underlying the mortality improvements model of CMI (2011).

- The experienced mortality improvement rates for the period 1991–2008 are in Tables C.1 to C.12.

- The age effects component of the 2008 mortality improvements are in Tables C.13 and C.14.

- The cohort component of the 2008 mortality improvements are in Tables C.15 and C.16.

- The long-term rates of mortality improvement for the age effects component are the same for males and females and are in Table C.17.

- The number of years taken to reach these long-term age effects component of improvement are in Table C.18.

- The long-term rate of improvement for the cohort component is zero.

- The number of years taken to reach the long-term rate of improvement for the cohort component is in Table C.19.

Table C.1: Experienced mortality improvement for males aged 20–59 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
20	0.47	0.35	0.25	0.23	0.32	0.51	0.81	1.19	1.66	2.20
21	0.22	0.20	0.17	0.15	0.22	0.40	0.68	1.06	1.50	2.00
22	-0.16	-0.08	-0.01	0.06	0.13	0.29	0.54	0.89	1.33	1.80
23	-0.60	-0.45	-0.30	-0.14	0.02	0.19	0.43	0.75	1.14	1.60
24	-1.04	-0.85	-0.66	-0.44	-0.21	0.04	0.32	0.64	1.00	1.41
25	-1.43	-1.21	-0.99	-0.76	-0.50	-0.21	0.13	0.50	0.88	1.28
26	-1.78	-1.52	-1.26	-1.02	-0.77	-0.48	-0.14	0.27	0.71	1.15
27	-2.06	-1.79	-1.49	-1.21	-0.95	-0.70	-0.39	-0.01	0.44	0.94
28	-2.08	-1.99	-1.69	-1.36	-1.07	-0.81	-0.56	-0.24	0.15	0.63
29	-1.87	-1.97	-1.83	-1.51	-1.18	-0.88	-0.63	-0.39	-0.09	0.32
30	-1.51	-1.74	-1.77	-1.60	-1.28	-0.96	-0.69	-0.46	-0.24	0.05
31	-1.10	-1.37	-1.52	-1.50	-1.33	-1.03	-0.75	-0.51	-0.31	-0.13
32	-0.76	-0.95	-1.14	-1.23	-1.19	-1.04	-0.78	-0.55	-0.35	-0.20
33	-0.50	-0.60	-0.72	-0.84	-0.90	-0.86	-0.74	-0.55	-0.37	-0.22
34	-0.31	-0.32	-0.36	-0.43	-0.51	-0.55	-0.53	-0.46	-0.33	-0.22
35	-0.21	-0.12	-0.08	-0.07	-0.11	-0.18	-0.22	-0.23	-0.21	-0.16
36	-0.22	-0.01	0.12	0.20	0.23	0.20	0.14	0.08	0.03	0.00
37	-0.38	-0.02	0.23	0.38	0.47	0.51	0.48	0.41	0.34	0.26
38	-0.73	-0.19	0.21	0.48	0.63	0.70	0.73	0.70	0.62	0.53
39	-1.18	-0.54	0.04	0.46	0.71	0.82	0.87	0.87	0.83	0.75
40	-1.32	-1.00	-0.31	0.29	0.70	0.92	0.98	0.98	0.95	0.90
41	-1.03	-1.16	-0.80	-0.09	0.53	0.93	1.11	1.12	1.07	1.02
42	-0.29	-0.89	-1.01	-0.63	0.11	0.73	1.12	1.28	1.25	1.17
43	0.81	-0.17	-0.79	-0.90	-0.51	0.25	0.89	1.28	1.44	1.40
44	1.77	0.94	-0.07	-0.70	-0.83	-0.43	0.35	1.01	1.42	1.59
45	2.43	1.91	1.06	0.02	-0.64	-0.78	-0.37	0.43	1.11	1.54
46	2.73	2.54	2.01	1.14	0.08	-0.60	-0.73	-0.32	0.50	1.21
47	2.69	2.78	2.60	2.07	1.20	0.12	-0.57	-0.69	-0.26	0.59
48	2.57	2.71	2.81	2.63	2.10	1.23	0.15	-0.52	-0.64	-0.18
49	2.45	2.58	2.73	2.83	2.65	2.13	1.27	0.20	-0.46	-0.55
50	2.30	2.43	2.58	2.74	2.85	2.68	2.18	1.33	0.29	-0.35
51	2.15	2.26	2.40	2.56	2.74	2.88	2.74	2.26	1.44	0.42
52	2.09	2.10	2.21	2.36	2.54	2.74	2.92	2.81	2.37	1.58
53	2.14	2.04	2.05	2.17	2.34	2.53	2.77	2.97	2.90	2.49
54	2.35	2.12	2.01	2.03	2.16	2.34	2.56	2.82	3.05	3.00
55	2.66	2.33	2.10	2.01	2.04	2.20	2.40	2.64	2.90	3.14
56	2.95	2.65	2.32	2.10	2.02	2.09	2.27	2.49	2.74	3.01
57	3.18	2.93	2.64	2.32	2.12	2.06	2.15	2.35	2.59	2.85
58	3.36	3.17	2.93	2.65	2.35	2.16	2.11	2.21	2.44	2.69
59	3.49	3.36	3.19	2.96	2.69	2.40	2.22	2.18	2.28	2.50

Table C.2: Experienced mortality improvement for males aged 20–59 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
20	2.79	3.31	3.72	4.02	4.23	4.39	4.52	4.62
21	2.54	3.10	3.57	3.92	4.14	4.26	4.35	4.41
22	2.31	2.84	3.37	3.79	4.08	4.23	4.27	4.29
23	2.08	2.58	3.09	3.58	3.96	4.20	4.29	4.26
24	1.86	2.33	2.80	3.28	3.74	4.08	4.27	4.32
25	1.69	2.11	2.55	2.99	3.43	3.86	4.17	4.32
26	1.56	1.96	2.34	2.73	3.13	3.55	3.95	4.22
27	1.41	1.83	2.20	2.54	2.88	3.25	3.63	4.00
28	1.17	1.66	2.07	2.42	2.71	3.02	3.34	3.69
29	0.81	1.37	1.87	2.28	2.61	2.87	3.14	3.44
30	0.45	0.96	1.53	2.05	2.46	2.78	3.02	3.27
31	0.15	0.54	1.05	1.64	2.17	2.59	2.92	3.17
32	-0.05	0.20	0.58	1.10	1.71	2.25	2.70	3.06
33	-0.12	-0.01	0.22	0.59	1.12	1.75	2.33	2.81
34	-0.13	-0.07	0.00	0.21	0.59	1.13	1.80	2.40
35	-0.11	-0.07	-0.05	0.00	0.20	0.59	1.16	1.86
36	-0.01	-0.03	-0.03	-0.03	0.00	0.20	0.61	1.21
37	0.18	0.11	0.05	0.01	0.00	0.03	0.24	0.66
38	0.43	0.32	0.21	0.13	0.08	0.07	0.12	0.35
39	0.65	0.55	0.43	0.31	0.23	0.19	0.20	0.28
40	0.82	0.73	0.62	0.51	0.41	0.34	0.34	0.38
41	0.94	0.86	0.77	0.68	0.59	0.51	0.48	0.51
42	1.09	0.99	0.89	0.80	0.72	0.66	0.62	0.63
43	1.30	1.20	1.08	0.95	0.85	0.78	0.74	0.74
44	1.55	1.46	1.34	1.20	1.04	0.92	0.85	0.83
45	1.74	1.73	1.64	1.52	1.35	1.16	1.02	0.95
46	1.67	1.89	1.91	1.84	1.72	1.53	1.31	1.15
47	1.32	1.81	2.06	2.11	2.05	1.93	1.73	1.50
48	0.69	1.45	1.97	2.24	2.30	2.26	2.14	1.95
49	-0.08	0.82	1.59	2.13	2.41	2.49	2.46	2.34
50	-0.43	0.05	0.96	1.75	2.28	2.57	2.65	2.62
51	-0.20	-0.27	0.22	1.13	1.91	2.44	2.71	2.77
52	0.59	-0.02	-0.08	0.41	1.30	2.06	2.57	2.83
53	1.73	0.76	0.17	0.12	0.59	1.47	2.21	2.70
54	2.61	1.87	0.93	0.35	0.29	0.76	1.61	2.33
55	3.10	2.72	1.99	1.06	0.48	0.42	0.88	1.71
56	3.25	3.20	2.82	2.10	1.17	0.60	0.53	0.97
57	3.12	3.35	3.30	2.92	2.20	1.27	0.70	0.62
58	2.95	3.22	3.44	3.39	3.01	2.30	1.39	0.81
59	2.76	3.02	3.29	3.51	3.46	3.10	2.41	1.52

Table C.3: Experienced mortality improvement for males aged 60–99 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
60	3.53	3.48	3.38	3.23	3.02	2.76	2.48	2.30	2.25	2.35
61	3.49	3.51	3.49	3.43	3.30	3.12	2.87	2.58	2.39	2.34
62	3.40	3.46	3.50	3.52	3.51	3.42	3.25	3.00	2.71	2.51
63	3.25	3.36	3.44	3.52	3.58	3.62	3.57	3.42	3.17	2.86
64	3.00	3.20	3.33	3.45	3.56	3.67	3.76	3.75	3.61	3.35
65	2.67	2.94	3.16	3.33	3.48	3.63	3.78	3.91	3.91	3.78
66	2.31	2.63	2.91	3.16	3.37	3.57	3.75	3.92	4.05	4.05
67	1.97	2.29	2.61	2.92	3.21	3.47	3.71	3.90	4.06	4.16
68	1.77	1.97	2.29	2.63	2.97	3.30	3.62	3.88	4.08	4.20
69	1.72	1.79	2.00	2.33	2.70	3.07	3.44	3.80	4.07	4.25
70	1.80	1.73	1.81	2.05	2.41	2.80	3.20	3.60	3.98	4.25
71	1.92	1.78	1.74	1.86	2.13	2.53	2.95	3.37	3.77	4.15
72	1.94	1.86	1.75	1.75	1.92	2.24	2.68	3.12	3.55	3.94
73	1.92	1.88	1.81	1.74	1.78	1.99	2.35	2.83	3.30	3.72
74	1.93	1.91	1.87	1.81	1.76	1.84	2.08	2.47	2.98	3.46
75	1.93	1.95	1.95	1.92	1.88	1.83	1.92	2.17	2.58	3.11
76	1.84	1.90	1.98	2.03	2.04	1.99	1.94	2.02	2.26	2.68
77	1.70	1.74	1.86	2.04	2.16	2.20	2.15	2.06	2.10	2.33
78	1.62	1.57	1.64	1.84	2.13	2.33	2.39	2.32	2.16	2.15
79	1.66	1.50	1.45	1.58	1.87	2.27	2.53	2.60	2.48	2.23
80	1.65	1.55	1.38	1.37	1.55	1.93	2.43	2.75	2.81	2.62
81	1.55	1.54	1.43	1.28	1.31	1.55	2.01	2.61	2.97	3.02
82	1.36	1.44	1.43	1.34	1.21	1.27	1.57	2.11	2.79	3.18
83	1.12	1.27	1.36	1.37	1.29	1.17	1.26	1.61	2.21	2.96
84	0.94	1.01	1.19	1.30	1.33	1.26	1.16	1.27	1.66	2.31
85	0.86	0.83	0.93	1.13	1.27	1.31	1.26	1.17	1.30	1.72
86	0.87	0.75	0.75	0.88	1.09	1.24	1.30	1.26	1.19	1.34
87	0.87	0.75	0.69	0.73	0.87	1.08	1.23	1.29	1.26	1.21
88	0.73	0.69	0.66	0.68	0.76	0.90	1.08	1.21	1.28	1.28
89	0.53	0.53	0.56	0.61	0.70	0.80	0.94	1.09	1.21	1.28
90	0.38	0.37	0.41	0.49	0.61	0.73	0.86	0.98	1.10	1.20
91	0.28	0.27	0.30	0.37	0.48	0.63	0.78	0.90	1.00	1.09
92	0.17	0.17	0.21	0.29	0.40	0.53	0.68	0.82	0.93	1.02
93	0.06	0.04	0.09	0.20	0.33	0.47	0.60	0.73	0.85	0.95
94	0.01	-0.07	-0.05	0.05	0.22	0.39	0.54	0.67	0.78	0.87
95	0.01	-0.09	-0.13	-0.07	0.06	0.27	0.45	0.60	0.71	0.80
96	-0.03	-0.09	-0.14	-0.13	-0.05	0.11	0.33	0.51	0.65	0.75
97	-0.11	-0.15	-0.17	-0.16	-0.10	0.01	0.19	0.40	0.58	0.71
98	-0.21	-0.24	-0.25	-0.22	-0.15	-0.05	0.09	0.27	0.48	0.65
99	-0.30	-0.35	-0.36	-0.32	-0.25	-0.13	0.01	0.18	0.36	0.56

Table C.4: Experienced mortality improvement for males aged 60–99 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
60	2.56	2.81	3.06	3.33	3.55	3.52	3.17	2.51
61	2.42	2.61	2.84	3.08	3.34	3.56	3.54	3.23
62	2.43	2.49	2.65	2.86	3.08	3.33	3.54	3.53
63	2.63	2.53	2.56	2.70	2.88	3.08	3.29	3.49
64	3.01	2.75	2.62	2.63	2.74	2.89	3.06	3.24
65	3.50	3.13	2.85	2.70	2.68	2.77	2.89	3.03
66	3.91	3.62	3.22	2.91	2.74	2.70	2.78	2.88
67	4.14	3.97	3.66	3.25	2.93	2.74	2.70	2.77
68	4.25	4.18	3.98	3.65	3.23	2.90	2.72	2.67
69	4.32	4.29	4.16	3.93	3.58	3.18	2.86	2.69
70	4.39	4.40	4.29	4.09	3.83	3.49	3.11	2.82
71	4.39	4.49	4.42	4.23	3.98	3.69	3.37	3.03
72	4.29	4.49	4.52	4.39	4.12	3.83	3.53	3.24
73	4.08	4.37	4.52	4.50	4.31	3.99	3.66	3.37
74	3.85	4.17	4.40	4.48	4.41	4.18	3.83	3.50
75	3.58	3.94	4.19	4.34	4.37	4.26	4.02	3.67
76	3.21	3.66	3.98	4.16	4.23	4.19	4.06	3.82
77	2.74	3.27	3.70	3.97	4.09	4.07	3.98	3.83
78	2.35	2.75	3.29	3.71	3.94	3.99	3.89	3.75
79	2.16	2.31	2.71	3.27	3.69	3.90	3.89	3.71
80	2.27	2.10	2.22	2.62	3.22	3.66	3.85	3.80
81	2.74	2.25	1.99	2.07	2.50	3.17	3.65	3.84
82	3.19	2.81	2.17	1.81	1.87	2.35	3.12	3.67
83	3.37	3.32	2.83	2.03	1.57	1.62	2.18	3.09
84	3.11	3.52	3.42	2.81	1.85	1.29	1.34	1.99
85	2.41	3.25	3.65	3.49	2.75	1.63	0.98	1.03
86	1.78	2.51	3.37	3.76	3.53	2.67	1.40	0.66
87	1.39	1.85	2.60	3.47	3.84	3.54	2.58	1.17
88	1.26	1.47	1.94	2.70	3.57	3.90	3.55	2.49
89	1.31	1.34	1.58	2.07	2.82	3.66	3.96	3.55
90	1.29	1.37	1.47	1.75	2.26	2.98	3.77	4.01
91	1.20	1.32	1.46	1.65	1.99	2.50	3.18	3.89
92	1.09	1.21	1.37	1.59	1.87	2.26	2.77	3.39
93	1.03	1.10	1.24	1.46	1.75	2.11	2.54	3.03
94	0.97	1.05	1.14	1.31	1.57	1.92	2.35	2.80
95	0.88	0.98	1.08	1.21	1.42	1.72	2.11	2.57
96	0.81	0.89	1.00	1.14	1.31	1.56	1.89	2.30
97	0.79	0.84	0.92	1.04	1.22	1.44	1.72	2.06
98	0.77	0.85	0.90	0.99	1.13	1.33	1.58	1.87
99	0.72	0.85	0.94	1.00	1.11	1.26	1.47	1.72

Table C.5: Experienced mortality improvement for males aged 100–120 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
100	-0.41	-0.45	-0.48	-0.45	-0.38	-0.26	-0.10	0.08	0.26	0.45
101	-0.31	-0.35	-0.38	-0.35	-0.28	-0.16	0.00	0.00	0.16	0.35
102	-0.21	-0.25	-0.28	-0.25	-0.18	-0.06	0.00	0.00	0.06	0.25
103	-0.11	-0.15	-0.18	-0.15	-0.08	0.00	0.00	0.00	0.00	0.15
104	-0.01	-0.05	-0.08	-0.05	0.00	0.00	0.00	0.00	0.00	0.05
105	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
106	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
107	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
109	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
110	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
111	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
113	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
114	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
115	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
116	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C.6: Experienced mortality improvement for males aged 100–12 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
100	0.64	0.81	0.95	1.05	1.14	1.26	1.43	1.63
101	0.54	0.71	0.85	0.95	1.04	1.16	1.33	1.53
102	0.44	0.61	0.75	0.85	0.94	1.06	1.23	1.43
103	0.34	0.51	0.65	0.75	0.84	0.96	1.13	1.33
104	0.24	0.41	0.55	0.65	0.74	0.86	1.03	1.23
105	0.14	0.31	0.45	0.55	0.64	0.76	0.93	1.13
106	0.04	0.21	0.35	0.45	0.54	0.66	0.83	1.03
107	0.00	0.11	0.25	0.35	0.44	0.56	0.73	0.93
108	0.00	0.01	0.15	0.25	0.34	0.46	0.63	0.83
109	0.00	0.00	0.05	0.15	0.24	0.36	0.53	0.73
110	0.00	0.00	0.00	0.05	0.14	0.26	0.43	0.63
111	0.00	0.00	0.00	0.00	0.04	0.16	0.33	0.53
112	0.00	0.00	0.00	0.00	0.00	0.06	0.23	0.43
113	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.33
114	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.23
115	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13
116	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C.7: Experienced mortality improvement for females aged 20–59 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
20	0.29	0.29	0.34	0.43	0.54	0.69	0.88	1.07	1.28	1.50
21	0.20	0.17	0.19	0.25	0.35	0.48	0.66	0.86	1.07	1.29
22	0.16	0.10	0.08	0.12	0.19	0.30	0.45	0.64	0.86	1.08
23	0.16	0.07	0.03	0.04	0.08	0.17	0.29	0.46	0.65	0.88
24	0.21	0.08	0.02	0.01	0.03	0.09	0.19	0.32	0.49	0.69
25	0.30	0.14	0.04	0.01	0.01	0.05	0.13	0.23	0.37	0.54
26	0.45	0.24	0.10	0.03	0.02	0.05	0.11	0.19	0.31	0.45
27	0.63	0.39	0.21	0.10	0.06	0.07	0.12	0.19	0.29	0.41
28	0.79	0.57	0.36	0.22	0.13	0.11	0.14	0.21	0.29	0.40
29	0.93	0.72	0.53	0.36	0.25	0.19	0.19	0.24	0.32	0.41
30	1.03	0.84	0.66	0.51	0.38	0.30	0.27	0.29	0.35	0.44
31	1.09	0.93	0.77	0.64	0.52	0.43	0.38	0.37	0.40	0.47
32	1.08	0.98	0.85	0.73	0.63	0.55	0.49	0.47	0.48	0.52
33	1.03	0.97	0.89	0.80	0.71	0.65	0.60	0.57	0.57	0.59
34	0.97	0.94	0.90	0.84	0.78	0.72	0.68	0.65	0.64	0.66
35	0.91	0.90	0.89	0.86	0.83	0.79	0.75	0.73	0.71	0.72
36	0.85	0.85	0.86	0.87	0.86	0.85	0.82	0.80	0.79	0.78
37	0.80	0.80	0.83	0.86	0.89	0.90	0.90	0.89	0.87	0.86
38	0.76	0.76	0.78	0.82	0.89	0.94	0.97	0.98	0.97	0.96
39	0.74	0.71	0.72	0.77	0.84	0.93	1.01	1.06	1.08	1.08
40	0.75	0.68	0.67	0.70	0.77	0.88	1.00	1.11	1.18	1.21
41	0.80	0.68	0.63	0.64	0.69	0.79	0.93	1.09	1.22	1.31
42	0.92	0.74	0.63	0.59	0.62	0.70	0.82	0.99	1.19	1.35
43	1.10	0.86	0.69	0.59	0.57	0.62	0.72	0.87	1.07	1.30
44	1.29	1.06	0.83	0.66	0.57	0.57	0.63	0.76	0.93	1.16
45	1.47	1.26	1.04	0.81	0.66	0.58	0.58	0.67	0.81	1.01
46	1.63	1.45	1.25	1.04	0.82	0.67	0.61	0.63	0.73	0.88
47	1.75	1.60	1.44	1.26	1.06	0.86	0.72	0.67	0.70	0.81
48	1.80	1.69	1.57	1.43	1.27	1.09	0.91	0.79	0.75	0.79
49	1.81	1.73	1.64	1.55	1.44	1.31	1.15	0.98	0.87	0.84
50	1.80	1.72	1.66	1.60	1.55	1.47	1.36	1.22	1.06	0.97
51	1.78	1.70	1.65	1.60	1.58	1.55	1.51	1.42	1.30	1.16
52	1.82	1.70	1.63	1.59	1.57	1.57	1.58	1.56	1.49	1.39
53	1.94	1.76	1.64	1.59	1.57	1.57	1.60	1.63	1.63	1.58
54	2.15	1.90	1.72	1.63	1.60	1.59	1.62	1.66	1.71	1.71
55	2.41	2.12	1.88	1.73	1.65	1.65	1.68	1.71	1.76	1.81
56	2.62	2.38	2.10	1.88	1.76	1.71	1.75	1.80	1.84	1.89
57	2.76	2.59	2.36	2.10	1.91	1.81	1.80	1.86	1.93	1.99
58	2.85	2.74	2.57	2.36	2.12	1.95	1.87	1.89	1.98	2.07
59	2.92	2.85	2.73	2.57	2.37	2.15	2.00	1.95	1.98	2.09

Table C.8: Experienced mortality improvement for females aged 20–59 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
20	1.72	1.91	2.08	2.21	2.30	2.37	2.41	2.41
21	1.51	1.74	1.93	2.08	2.19	2.27	2.31	2.33
22	1.31	1.53	1.75	1.93	2.07	2.17	2.22	2.25
23	1.11	1.33	1.55	1.76	1.93	2.05	2.13	2.17
24	0.91	1.14	1.36	1.57	1.76	1.92	2.03	2.09
25	0.74	0.96	1.18	1.39	1.59	1.76	1.90	2.00
26	0.62	0.81	1.03	1.24	1.43	1.61	1.77	1.90
27	0.55	0.72	0.90	1.10	1.30	1.48	1.64	1.79
28	0.52	0.67	0.83	1.00	1.19	1.37	1.53	1.68
29	0.52	0.65	0.79	0.95	1.11	1.29	1.45	1.60
30	0.54	0.65	0.78	0.92	1.07	1.23	1.39	1.54
31	0.56	0.67	0.78	0.91	1.05	1.19	1.35	1.51
32	0.60	0.69	0.80	0.91	1.04	1.17	1.32	1.47
33	0.65	0.73	0.82	0.92	1.04	1.16	1.30	1.44
34	0.70	0.76	0.84	0.94	1.05	1.17	1.29	1.43
35	0.74	0.79	0.86	0.95	1.06	1.17	1.29	1.42
36	0.78	0.81	0.87	0.95	1.05	1.17	1.29	1.41
37	0.86	0.86	0.88	0.94	1.03	1.15	1.27	1.40
38	0.95	0.94	0.94	0.96	1.02	1.12	1.25	1.38
39	1.07	1.06	1.04	1.03	1.05	1.11	1.21	1.35
40	1.20	1.19	1.18	1.15	1.13	1.15	1.21	1.32
41	1.34	1.34	1.33	1.30	1.27	1.25	1.26	1.32
42	1.45	1.49	1.49	1.47	1.44	1.41	1.37	1.38
43	1.48	1.60	1.64	1.63	1.61	1.58	1.54	1.51
44	1.41	1.62	1.74	1.79	1.78	1.75	1.72	1.68
45	1.25	1.53	1.75	1.88	1.94	1.92	1.89	1.86
46	1.09	1.35	1.64	1.86	2.01	2.06	2.05	2.02
47	0.97	1.18	1.45	1.74	1.97	2.11	2.17	2.16
48	0.90	1.07	1.28	1.54	1.83	2.05	2.19	2.26
49	0.89	1.00	1.17	1.38	1.63	1.90	2.12	2.25
50	0.94	0.99	1.10	1.26	1.46	1.69	1.95	2.16
51	1.07	1.04	1.09	1.19	1.34	1.52	1.74	1.98
52	1.25	1.16	1.14	1.17	1.27	1.40	1.57	1.77
53	1.48	1.34	1.25	1.22	1.25	1.33	1.45	1.61
54	1.67	1.57	1.43	1.33	1.29	1.31	1.39	1.50
55	1.81	1.76	1.65	1.50	1.39	1.35	1.36	1.43
56	1.92	1.91	1.85	1.73	1.57	1.45	1.40	1.41
57	2.03	2.05	2.02	1.94	1.80	1.63	1.51	1.45
58	2.13	2.16	2.16	2.12	2.03	1.88	1.70	1.57
59	2.18	2.24	2.27	2.26	2.21	2.11	1.96	1.78

Table C.9: Experienced mortality improvement for females aged 60–99 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
60	2.95	2.96	2.88	2.75	2.59	2.40	2.20	2.07	2.02	2.07
61	2.92	3.02	3.02	2.92	2.78	2.63	2.45	2.26	2.15	2.11
62	2.78	2.99	3.09	3.08	2.97	2.83	2.67	2.50	2.34	2.23
63	2.51	2.83	3.05	3.16	3.14	3.02	2.88	2.73	2.57	2.41
64	2.16	2.53	2.87	3.10	3.22	3.20	3.09	2.94	2.80	2.64
65	1.79	2.17	2.55	2.91	3.16	3.28	3.27	3.15	3.01	2.86
66	1.44	1.81	2.19	2.59	2.97	3.24	3.36	3.35	3.23	3.07
67	1.13	1.46	1.83	2.23	2.64	3.05	3.34	3.47	3.45	3.31
68	0.92	1.16	1.49	1.87	2.28	2.72	3.15	3.45	3.58	3.55
69	0.84	0.96	1.20	1.54	1.93	2.35	2.80	3.25	3.56	3.69
70	0.87	0.87	1.00	1.24	1.60	2.00	2.43	2.90	3.35	3.66
71	0.98	0.89	0.90	1.04	1.30	1.68	2.09	2.53	2.99	3.44
72	1.05	0.98	0.91	0.93	1.09	1.38	1.77	2.20	2.63	3.07
73	1.08	1.04	0.99	0.93	0.98	1.15	1.46	1.87	2.30	2.72
74	1.10	1.08	1.05	1.01	0.97	1.04	1.23	1.55	1.96	2.38
75	1.13	1.11	1.10	1.08	1.06	1.04	1.13	1.33	1.64	2.04
76	1.14	1.13	1.13	1.14	1.13	1.13	1.14	1.23	1.43	1.73
77	1.14	1.11	1.12	1.16	1.19	1.22	1.23	1.25	1.34	1.54
78	1.15	1.09	1.08	1.12	1.21	1.28	1.33	1.35	1.37	1.46
79	1.17	1.07	1.03	1.05	1.14	1.28	1.40	1.47	1.49	1.50
80	1.16	1.07	0.98	0.97	1.03	1.18	1.38	1.54	1.63	1.65
81	1.12	1.04	0.96	0.89	0.91	1.02	1.23	1.50	1.70	1.80
82	1.08	0.99	0.92	0.85	0.81	0.86	1.02	1.29	1.62	1.86
83	1.06	0.96	0.88	0.81	0.76	0.75	0.84	1.04	1.35	1.74
84	1.06	0.94	0.85	0.78	0.73	0.70	0.72	0.83	1.07	1.43
85	1.05	0.91	0.81	0.74	0.70	0.67	0.67	0.71	0.85	1.12
86	1.04	0.88	0.77	0.70	0.66	0.65	0.66	0.68	0.74	0.90
87	1.03	0.85	0.71	0.62	0.59	0.61	0.64	0.68	0.73	0.80
88	0.96	0.80	0.64	0.53	0.49	0.50	0.58	0.65	0.73	0.81
89	0.88	0.73	0.58	0.45	0.38	0.37	0.44	0.56	0.69	0.80
90	0.82	0.65	0.51	0.38	0.29	0.26	0.30	0.40	0.57	0.73
91	0.77	0.60	0.44	0.32	0.23	0.18	0.19	0.26	0.39	0.58
92	0.70	0.56	0.40	0.27	0.18	0.13	0.12	0.16	0.26	0.41
93	0.59	0.49	0.37	0.24	0.14	0.09	0.07	0.10	0.18	0.30
94	0.46	0.39	0.30	0.21	0.11	0.05	0.03	0.06	0.13	0.23
95	0.32	0.27	0.21	0.14	0.08	0.02	0.00	0.02	0.08	0.18
96	0.18	0.14	0.10	0.06	0.02	-0.01	-0.02	-0.01	0.04	0.13
97	0.05	0.01	-0.02	-0.04	-0.06	-0.07	-0.06	-0.04	0.01	0.09
98	-0.02	-0.08	-0.12	-0.14	-0.15	-0.14	-0.12	-0.08	-0.03	0.05
99	-0.04	-0.12	-0.18	-0.22	-0.23	-0.22	-0.19	-0.14	-0.08	0.00

Table C.10: Experienced mortality improvement for females aged 60–99 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
60	2.18	2.27	2.33	2.36	2.35	2.29	2.19	2.04
61	2.15	2.26	2.35	2.40	2.43	2.41	2.35	2.25
62	2.19	2.24	2.33	2.41	2.46	2.48	2.45	2.40
63	2.31	2.28	2.32	2.40	2.47	2.51	2.51	2.48
64	2.49	2.40	2.36	2.39	2.46	2.52	2.54	2.54
65	2.71	2.57	2.47	2.43	2.45	2.52	2.56	2.57
66	2.92	2.77	2.62	2.52	2.48	2.50	2.55	2.59
67	3.14	2.97	2.81	2.66	2.55	2.51	2.52	2.58
68	3.39	3.20	3.02	2.84	2.68	2.57	2.52	2.53
69	3.65	3.46	3.25	3.05	2.87	2.70	2.59	2.53
70	3.79	3.73	3.53	3.30	3.09	2.90	2.73	2.61
71	3.73	3.85	3.78	3.57	3.34	3.13	2.94	2.77
72	3.50	3.78	3.88	3.81	3.60	3.38	3.17	2.99
73	3.14	3.54	3.79	3.88	3.82	3.61	3.40	3.21
74	2.79	3.19	3.55	3.78	3.86	3.80	3.61	3.41
75	2.45	2.84	3.21	3.53	3.75	3.82	3.76	3.59
76	2.12	2.50	2.86	3.20	3.50	3.69	3.76	3.71
77	1.82	2.18	2.54	2.87	3.18	3.45	3.63	3.69
78	1.64	1.91	2.24	2.57	2.88	3.16	3.40	3.55
79	1.57	1.73	1.98	2.30	2.61	2.89	3.13	3.35
80	1.62	1.67	1.80	2.04	2.35	2.64	2.90	3.12
81	1.80	1.73	1.74	1.86	2.08	2.39	2.68	2.92
82	1.96	1.94	1.82	1.79	1.89	2.11	2.42	2.71
83	2.00	2.11	2.05	1.88	1.81	1.88	2.11	2.44
84	1.85	2.13	2.23	2.14	1.91	1.80	1.86	2.10
85	1.50	1.95	2.25	2.33	2.20	1.93	1.78	1.83
86	1.18	1.58	2.05	2.34	2.41	2.25	1.94	1.77
87	0.97	1.26	1.66	2.12	2.41	2.46	2.29	1.95
88	0.90	1.07	1.36	1.75	2.20	2.47	2.51	2.32
89	0.90	1.01	1.19	1.48	1.85	2.27	2.52	2.55
90	0.87	1.00	1.13	1.33	1.60	1.96	2.35	2.58
91	0.77	0.94	1.10	1.26	1.47	1.74	2.09	2.45
92	0.61	0.81	1.00	1.19	1.38	1.60	1.88	2.20
93	0.46	0.66	0.87	1.07	1.27	1.48	1.72	1.99
94	0.37	0.54	0.74	0.94	1.15	1.36	1.58	1.81
95	0.31	0.46	0.64	0.84	1.04	1.24	1.44	1.65
96	0.26	0.40	0.57	0.76	0.96	1.15	1.34	1.52
97	0.21	0.35	0.51	0.68	0.87	1.07	1.26	1.43
98	0.16	0.29	0.45	0.62	0.79	0.98	1.17	1.35
99	0.11	0.23	0.38	0.54	0.71	0.89	1.07	1.25

Table C.11: Experienced mortality improvement for males aged 100–120 in the years 1991–2000 (%). Source: CMI (2011).

Age	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
100	-0.06	-0.13	-0.21	-0.26	-0.29	-0.29	-0.26	-0.21	-0.14	-0.05
101	0.00	-0.03	-0.11	-0.16	-0.19	-0.19	-0.16	-0.11	-0.04	0.00
102	0.00	0.00	-0.01	-0.06	-0.09	-0.09	-0.06	-0.01	0.00	0.00
103	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
104	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
105	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
106	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
107	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
109	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
110	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
111	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
113	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
114	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
115	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
116	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C.12: Experienced mortality improvement for males aged 100–12 in the years 2001–2008 (%). Source: CMI (2011).

Age	2001	2002	2003	2004	2005	2006	2007	2008
100	0.05	0.17	0.31	0.46	0.63	0.80	0.97	1.14
101	0.00	0.07	0.21	0.36	0.53	0.70	0.87	1.04
102	0.00	0.00	0.11	0.26	0.43	0.60	0.77	0.94
103	0.00	0.00	0.01	0.16	0.33	0.50	0.67	0.84
104	0.00	0.00	0.00	0.06	0.23	0.40	0.57	0.74
105	0.00	0.00	0.00	0.00	0.13	0.30	0.47	0.64
106	0.00	0.00	0.00	0.00	0.03	0.20	0.37	0.54
107	0.00	0.00	0.00	0.00	0.00	0.10	0.27	0.44
108	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.34
109	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.24
110	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14
111	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
113	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
114	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
115	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
116	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C.13: Age effects component of 2008 mortality improvement for males (%). Source CMI (2011).

Age	Improvement	Age	Improvement	Age	Improvement
20	4.62	60	3.01	100	2.08
21	4.41	61	3.04	101	1.96
22	4.29	62	3.05	102	1.83
23	4.26	63	3.06	103	1.70
24	4.32	64	3.05	104	1.57
25	4.32	65	3.05	105	1.45
26	4.22	66	3.05	106	1.32
27	4.00	67	3.05	107	1.19
28	3.69	68	3.06	108	1.06
29	3.44	69	3.08	109	0.94
30	3.29	70	3.10	110	0.81
31	3.22	71	3.12	111	0.68
32	3.18	72	3.13	112	0.55
33	3.08	73	3.14	113	0.42
34	2.89	74	3.13	114	0.30
35	2.61	75	3.11	115	0.17
36	2.26	76	3.08	116	0.04
37	2.01	77	3.04	117	0.00
38	1.92	78	3.01	118	0.00
39	1.99	79	3.00	119	0.00
40	2.16	80	2.98	120	0.00
41	2.31	81	2.96		
42	2.42	82	2.92		
43	2.49	83	2.85		
44	2.53	84	2.77		
45	2.56	85	2.69		
46	2.61	86	2.63		
47	2.68	87	2.60		
48	2.76	88	2.59		
49	2.78	89	2.58		
50	2.77	90	2.56		
51	2.77	91	2.53		
52	2.81	92	2.48		
53	2.87	93	2.42		
54	2.91	94	2.35		
55	2.90	95	2.29		
56	2.86	96	2.23		
57	2.86	97	2.17		
58	2.90	98	2.13		
59	2.96	99	2.10		

Table C.14: Age effects component of 2008 mortality improvement for females (%). Source CMI (2011).

Age	Improvement	Age	Improvement	Age	Improvement
20	2.41	60	2.01	100	1.39
21	2.33	61	2.05	101	1.26
22	2.25	62	2.10	102	1.14
23	2.17	63	2.13	103	1.02
24	2.09	64	2.16	104	0.90
25	2.00	65	2.19	105	0.78
26	1.90	66	2.23	106	0.66
27	1.79	67	2.27	107	0.53
28	1.68	68	2.33	108	0.41
29	1.60	69	2.39	109	0.29
30	1.56	70	2.45	110	0.17
31	1.57	71	2.51	111	0.05
32	1.61	72	2.57	112	0.00
33	1.70	73	2.62	113	0.00
34	1.81	74	2.65	114	0.00
35	1.91	75	2.67	115	0.00
36	1.98	76	2.68	116	0.00
37	1.99	77	2.68	117	0.00
38	1.95	78	2.68	118	0.00
39	1.89	79	2.68	119	0.00
40	1.82	80	2.69	120	0.00
41	1.79	81	2.69		
42	1.78	82	2.68		
43	1.80	83	2.64		
44	1.83	84	2.58		
45	1.84	85	2.51		
46	1.83	86	2.44		
47	1.81	87	2.38		
48	1.79	88	2.31		
49	1.77	89	2.24		
50	1.73	90	2.17		
51	1.70	91	2.09		
52	1.68	92	2.00		
53	1.68	93	1.92		
54	1.71	94	1.84		
55	1.75	95	1.76		
56	1.79	96	1.68		
57	1.83	97	1.61		
58	1.89	98	1.54		
59	1.95	99	1.46		

Table C.15: Cohort component of 2008 mortality improvement for males (%). Source CMI (2011).

Birth Year	Improvement	Birth Year	Improvement	Birth Year	Improvement
1988	0.00	1948	-0.50	1908	-0.45
1987	0.00	1947	0.19	1907	-0.42
1986	0.00	1946	0.48	1906	-0.40
1985	0.00	1945	0.43	1905	-0.37
1984	0.00	1944	0.19	1904	-0.34
1983	0.00	1943	-0.02	1903	-0.31
1982	0.00	1942	-0.16	1902	-0.29
1981	0.00	1941	-0.29	1901	-0.26
1980	0.00	1940	-0.39	1900	-0.23
1979	0.00	1939	-0.40	1899	-0.20
1978	-0.02	1938	-0.28	1898	-0.17
1977	-0.05	1937	-0.09	1897	-0.15
1976	-0.13	1936	0.10	1896	-0.12
1975	-0.27	1935	0.23	1895	-0.09
1974	-0.49	1934	0.37	1894	-0.06
1973	-0.75	1933	0.56	1893	-0.04
1972	-1.05	1932	0.75	1892	-0.01
1971	-1.34	1931	0.79	1891	0.00
1970	-1.57	1930	0.73	1890	0.00
1969	-1.71	1929	0.71	1889	0.00
1968	-1.77	1928	0.82	1888	0.00
1967	-1.79	1927	0.88		
1966	-1.79	1926	0.75		
1965	-1.75	1925	0.24		
1964	-1.70	1924	-0.78		
1963	-1.61	1923	-1.66		
1962	-1.45	1922	-1.97		
1961	-1.18	1921	-1.43		
1960	-0.81	1920	-0.10		
1959	-0.44	1919	0.97		
1958	-0.15	1918	1.45		
1957	0.01	1917	1.36		
1956	0.02	1916	0.91		
1955	-0.17	1915	0.61		
1954	-0.58	1914	0.45		
1953	-1.19	1913	0.29		
1952	-1.89	1912	0.07		
1951	-2.23	1911	-0.12		
1950	-2.09	1910	-0.26		
1949	-1.45	1909	-0.38		

Table C.16: Cohort component of 2008 mortality improvement for females (%). Source CMI (2011).

Birth Year	Improvement	Birth Year	Improvement	Birth Year	Improvement
1988	0.00	1948	0.04	1908	-0.25
1987	0.00	1947	0.20	1907	-0.22
1986	0.00	1946	0.30	1906	-0.20
1985	0.00	1945	0.35	1905	-0.18
1984	0.00	1944	0.38	1904	-0.16
1983	0.00	1943	0.38	1903	-0.14
1982	0.00	1942	0.37	1902	-0.12
1981	0.00	1941	0.31	1901	-0.09
1980	0.00	1940	0.21	1900	-0.07
1979	0.00	1939	0.15	1899	-0.05
1978	-0.02	1938	0.16	1898	-0.03
1977	-0.06	1937	0.26	1897	-0.01
1976	-0.14	1936	0.42	1896	0.00
1975	-0.26	1935	0.59	1895	0.00
1974	-0.38	1934	0.76	1894	0.00
1973	-0.49	1933	0.92	1893	0.00
1972	-0.56	1932	1.03	1892	0.00
1971	-0.58	1931	1.02	1891	0.00
1970	-0.57	1930	0.88	1890	0.00
1969	-0.54	1929	0.66	1889	0.00
1968	-0.51	1928	0.43	1888	0.00
1967	-0.47	1927	0.23		
1966	-0.40	1926	0.04		
1965	-0.30	1925	-0.19		
1964	-0.15	1924	-0.49		
1963	0.02	1923	-0.68		
1962	0.19	1922	-0.68		
1961	0.35	1921	-0.42		
1960	0.46	1920	0.01		
1959	0.49	1919	0.31		
1958	0.42	1918	0.42		
1957	0.28	1917	0.36		
1956	0.10	1916	0.20		
1955	-0.08	1915	0.07		
1954	-0.21	1914	-0.03		
1953	-0.32	1913	-0.11		
1952	-0.38	1912	-0.16		
1951	-0.39	1911	-0.18		
1950	-0.32	1910	-0.19		
1949	-0.17	1909	-0.21		

Table C.17: Long-term rate of age effects component of mortality improvement (%). Source CMI (2011).

Age	Improvement	Age	Improvement	Age	Improvement
20	1.00	60	1.00	100	0.67
21	1.00	61	1.00	101	0.63
22	1.00	62	1.00	102	0.60
23	1.00	63	1.00	103	0.57
24	1.00	64	1.00	104	0.53
25	1.00	65	1.00	105	0.50
26	1.00	66	1.00	106	0.47
27	1.00	67	1.00	107	0.43
28	1.00	68	1.00	108	0.40
29	1.00	69	1.00	109	0.37
30	1.00	70	1.00	110	0.33
31	1.00	71	1.00	111	0.30
32	1.00	72	1.00	112	0.27
33	1.00	73	1.00	113	0.23
34	1.00	74	1.00	114	0.20
35	1.00	75	1.00	115	0.17
36	1.00	76	1.00	116	0.13
37	1.00	77	1.00	117	0.10
38	1.00	78	1.00	118	0.07
39	1.00	79	1.00	119	0.03
40	1.00	80	1.00	120	0.00
41	1.00	81	1.00		
42	1.00	82	1.00		
43	1.00	83	1.00		
44	1.00	84	1.00		
45	1.00	85	1.00		
46	1.00	86	1.00		
47	1.00	87	1.00		
48	1.00	88	1.00		
49	1.00	89	1.00		
50	1.00	90	1.00		
51	1.00	91	0.97		
52	1.00	92	0.93		
53	1.00	93	0.90		
54	1.00	94	0.87		
55	1.00	95	0.83		
56	1.00	96	0.80		
57	1.00	97	0.77		
58	1.00	98	0.73		
59	1.00	99	0.70		

Table C.18: Number of years to reach long-term improvement rate for age effects component. Source CMI (2011).

Age	Years	Age	Years	Age	Years
20	10	60	20	100	5
21	10	61	20	101	5
22	10	62	20	102	5
23	10	63	20	103	5
24	10	64	20	104	5
25	10	65	20	105	5
26	10	66	20	106	5
27	10	67	20	107	5
28	10	68	20	108	5
29	10	69	20	109	5
30	10	70	20	110	5
31	10	71	20	111	5
32	10	72	20	112	5
33	10	73	20	113	5
34	10	74	20	114	5
35	10	75	20	115	5
36	10	76	20	116	5
37	10	77	20	117	5
38	10	78	20	118	5
39	10	79	20	119	5
40	10	80	20	120	5
41	10	81	19		
42	10	82	18		
43	10	83	17		
44	10	84	16		
45	10	85	15		
46	10	86	14		
47	10	87	13		
48	10	88	12		
49	10	89	11		
50	10	90	10		
51	11	91	9		
52	12	92	8		
53	13	93	7		
54	14	94	6		
55	15	95	5		
56	16	96	5		
57	17	97	5		
58	18	98	5		
59	19	99	5		

Table C.19: Number of years to reach long-term improvement rate for cohort component. Source CMI (2011).

| Birth Year	Years	Birth Year	Years	Birth Year	Years
1988	40	1948	40	1908	5
1987	40	1947	39	1907	5
1986	40	1946	38	1906	5
1985	40	1945	37	1905	5
1984	40	1944	36	1904	5
1983	40	1943	35	1903	5
1982	40	1942	34	1902	5
1981	40	1941	33	1901	5
1980	40	1940	32	1900	5
1979	40	1939	31	1899	5
1978	40	1938	30	1898	5
1977	40	1937	29	1897	5
1976	40	1936	28	1896	5
1975	40	1935	27	1895	5
1974	40	1934	26	1894	5
1973	40	1933	25	1893	5
1972	40	1932	24	1892	5
1971	40	1931	23	1891	5
1970	40	1930	22	1890	5
1969	40	1929	21	1889	5
1968	40	1928	20	1888	5
1967	40	1927	19		
1966	40	1926	18		
1965	40	1925	17		
1964	40	1924	16		
1963	40	1923	15		
1962	40	1922	14		
1961	40	1921	13		
1960	40	1920	12		
1959	40	1919	11		
1958	40	1918	10		
1957	40	1917	9		
1956	40	1916	8		
1955	40	1915	7		
1954	40	1914	6		
1953	40	1913	5		
1952	40	1912	5		
1951	40	1911	5		
1950	40	1910	5		
1949	40	1909	5		

Appendix D

LTC Single Premium Rates
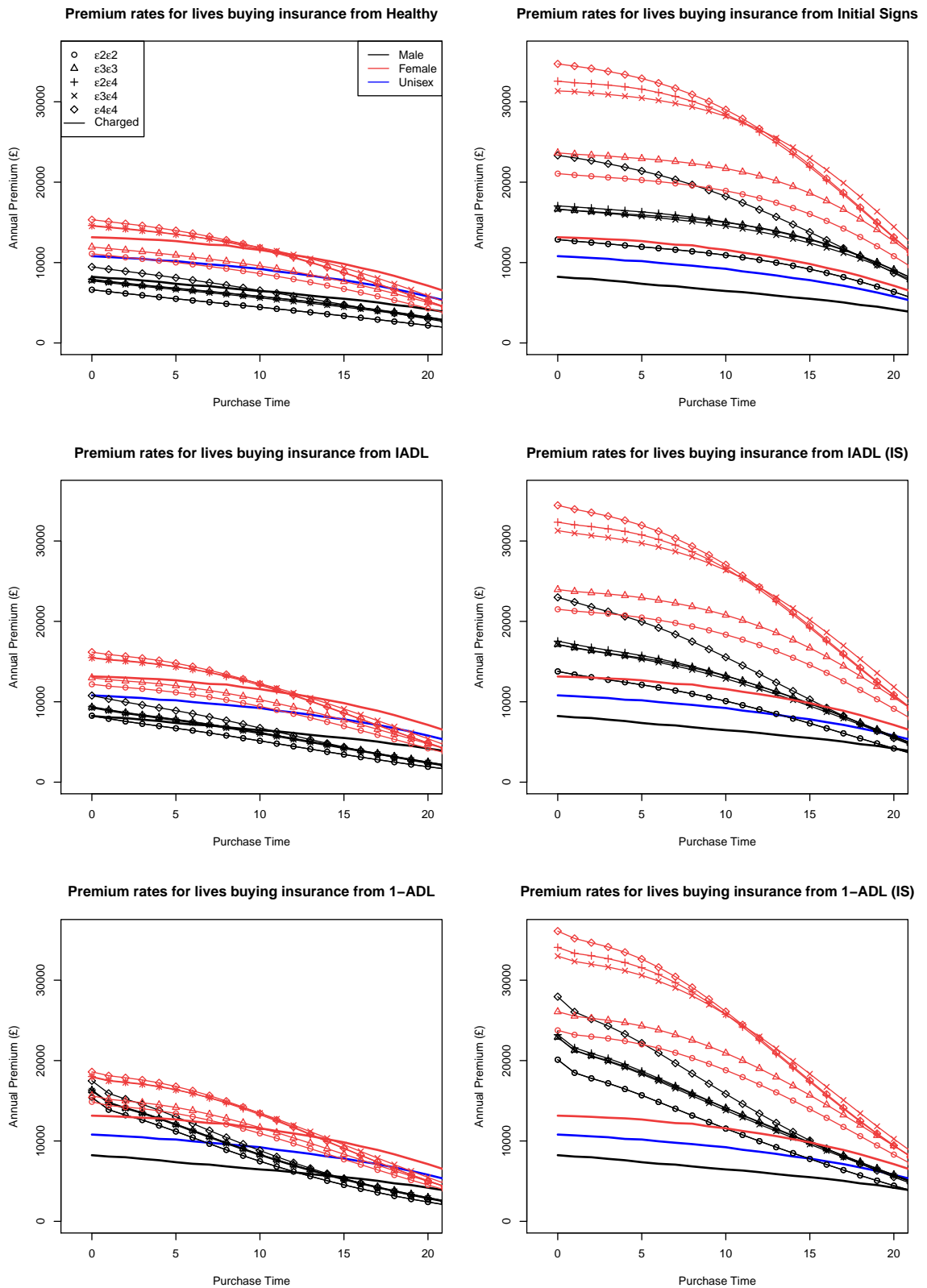
Figure D.1: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 62.5 at 1st January, 2013. The same legend is used throughout the plots.
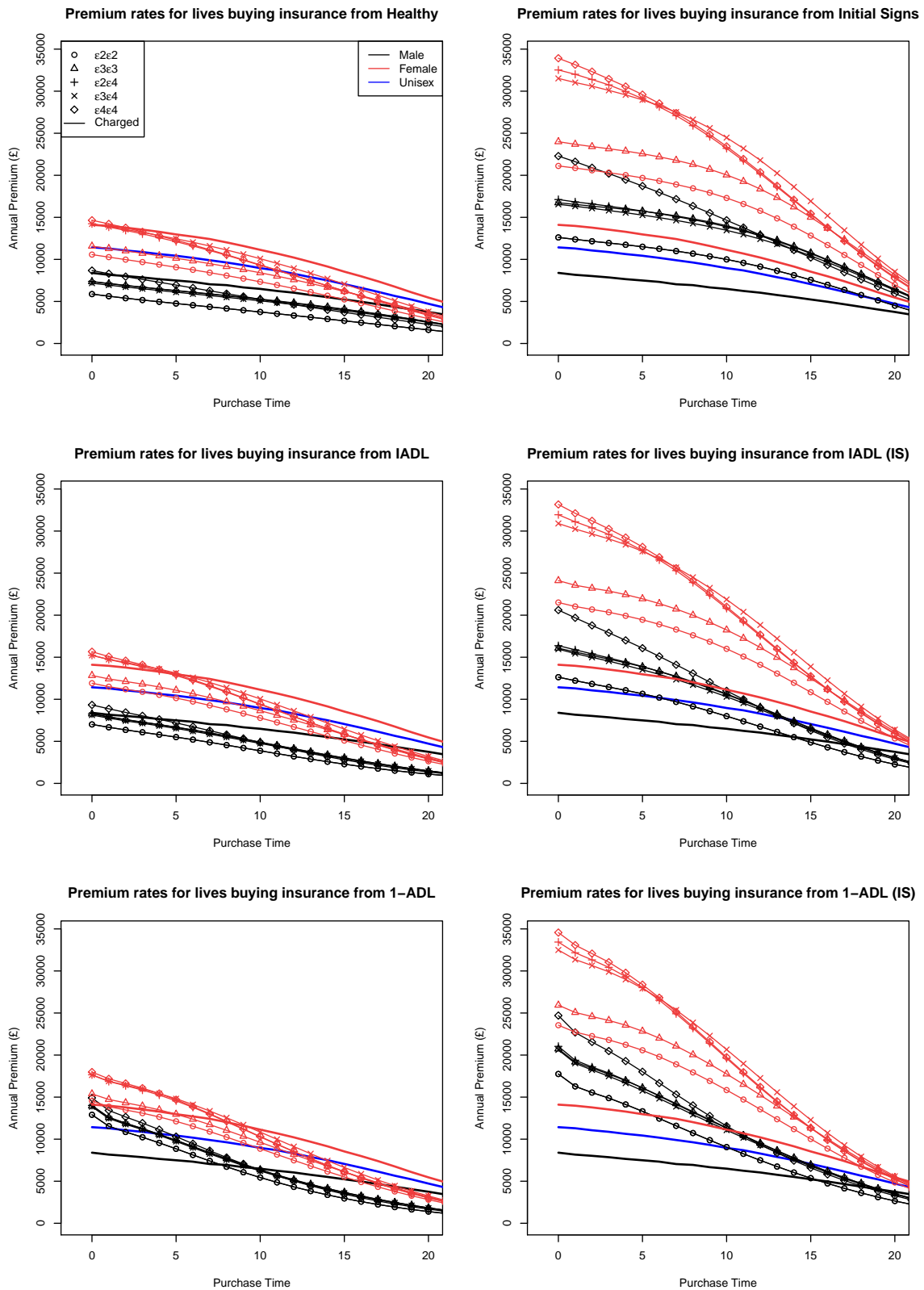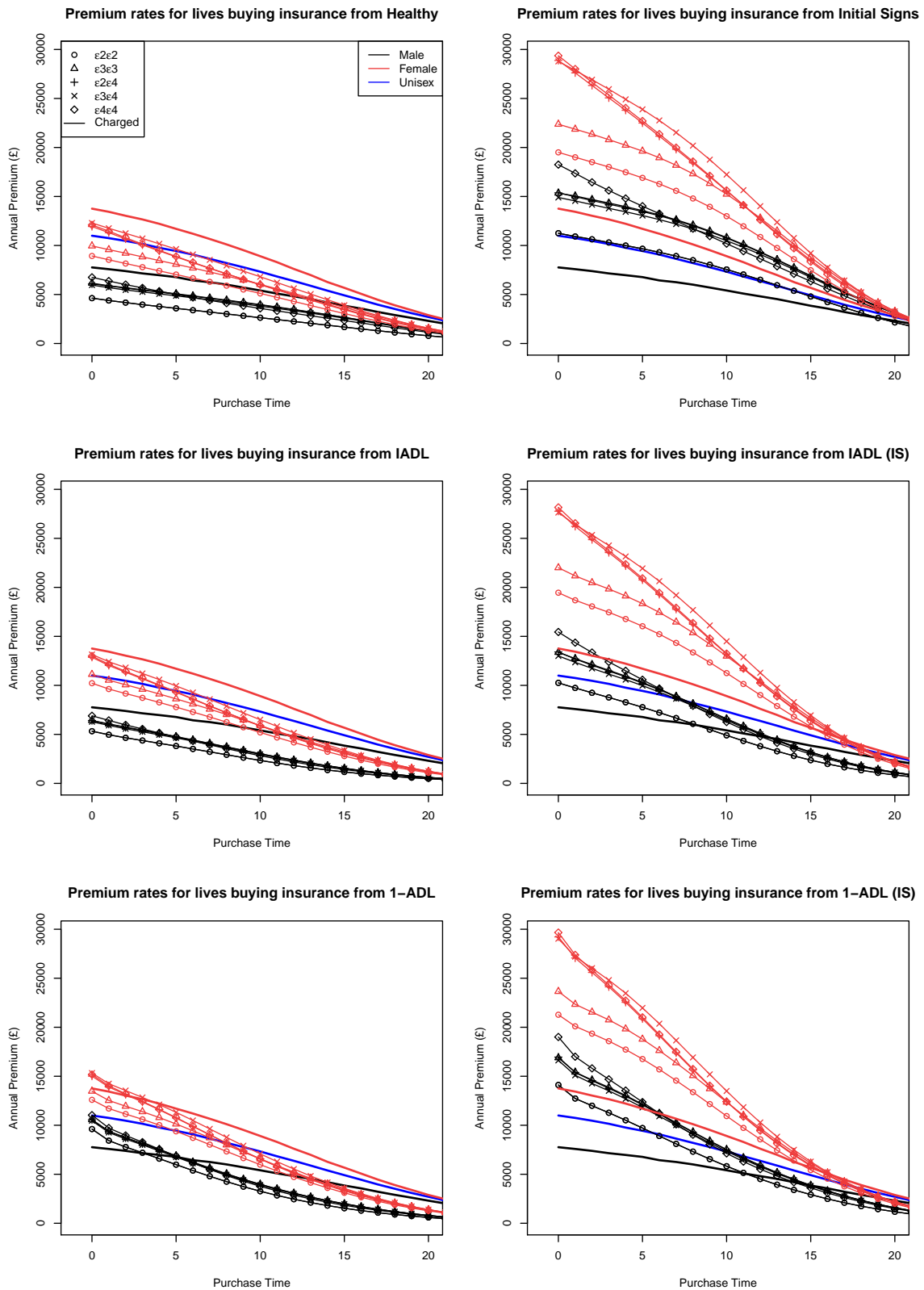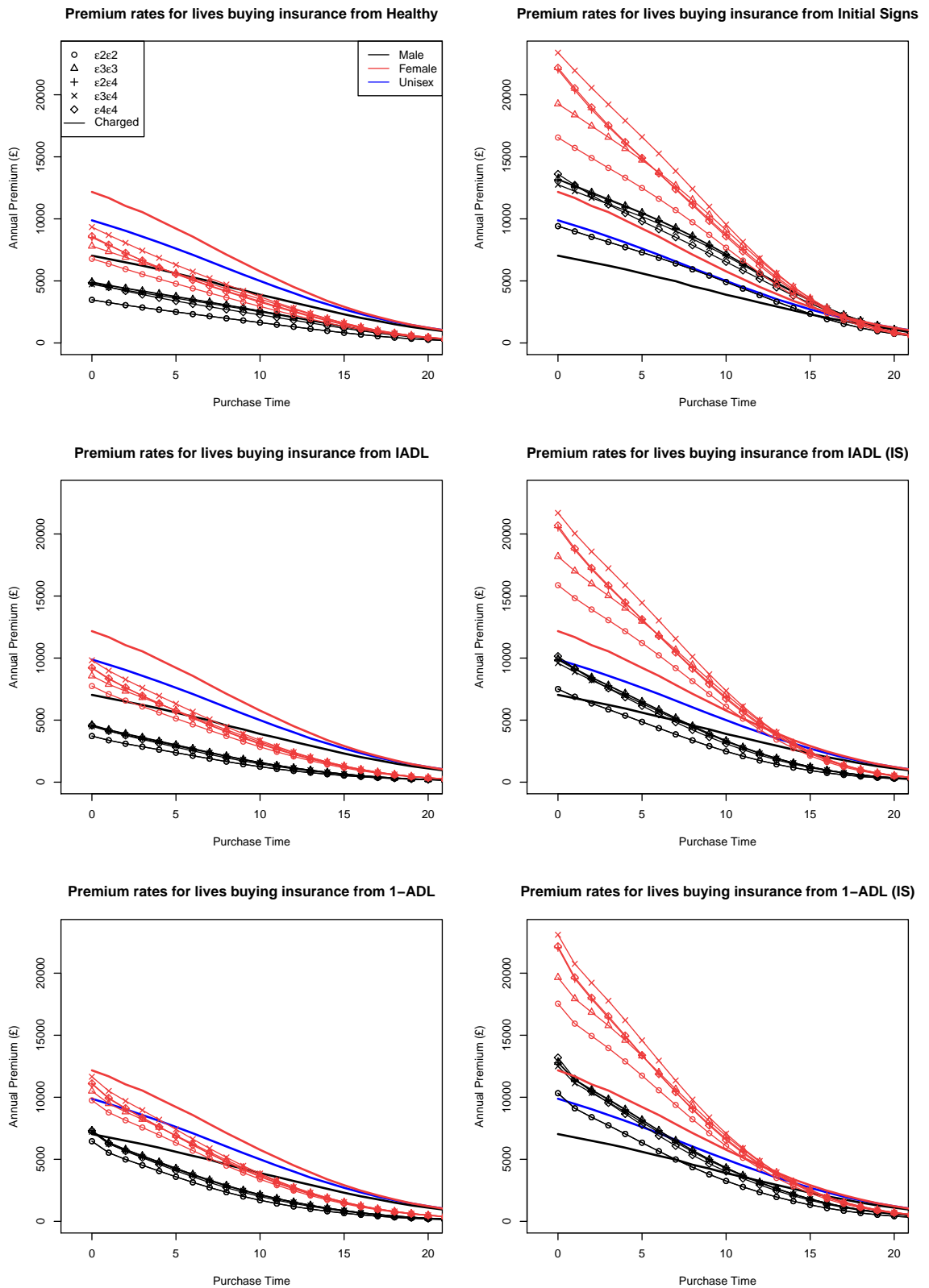
Figure D.2: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 67.5 at 1st January, 2013. The same legend is used throughout the plots.

Figure D.3: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 72.5 at 1st January, 2013. The same legend is used throughout the plots.
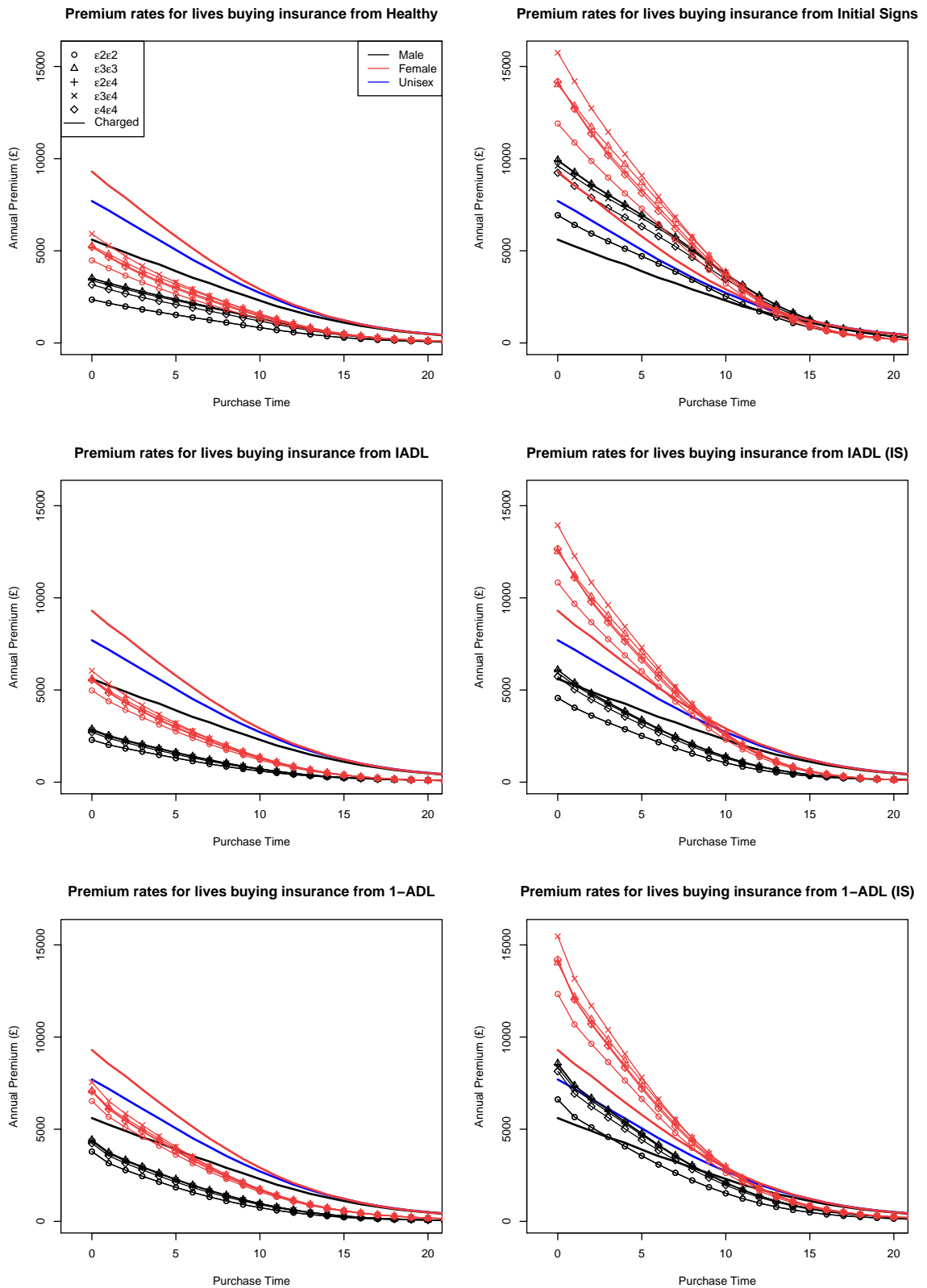
Figure D.4: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 77.5 at 1st January, 2013. The same legend is used throughout the plots.

Figure D.5: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 82.5 at 1st January, 2013. The same legend is used throughout the plots.
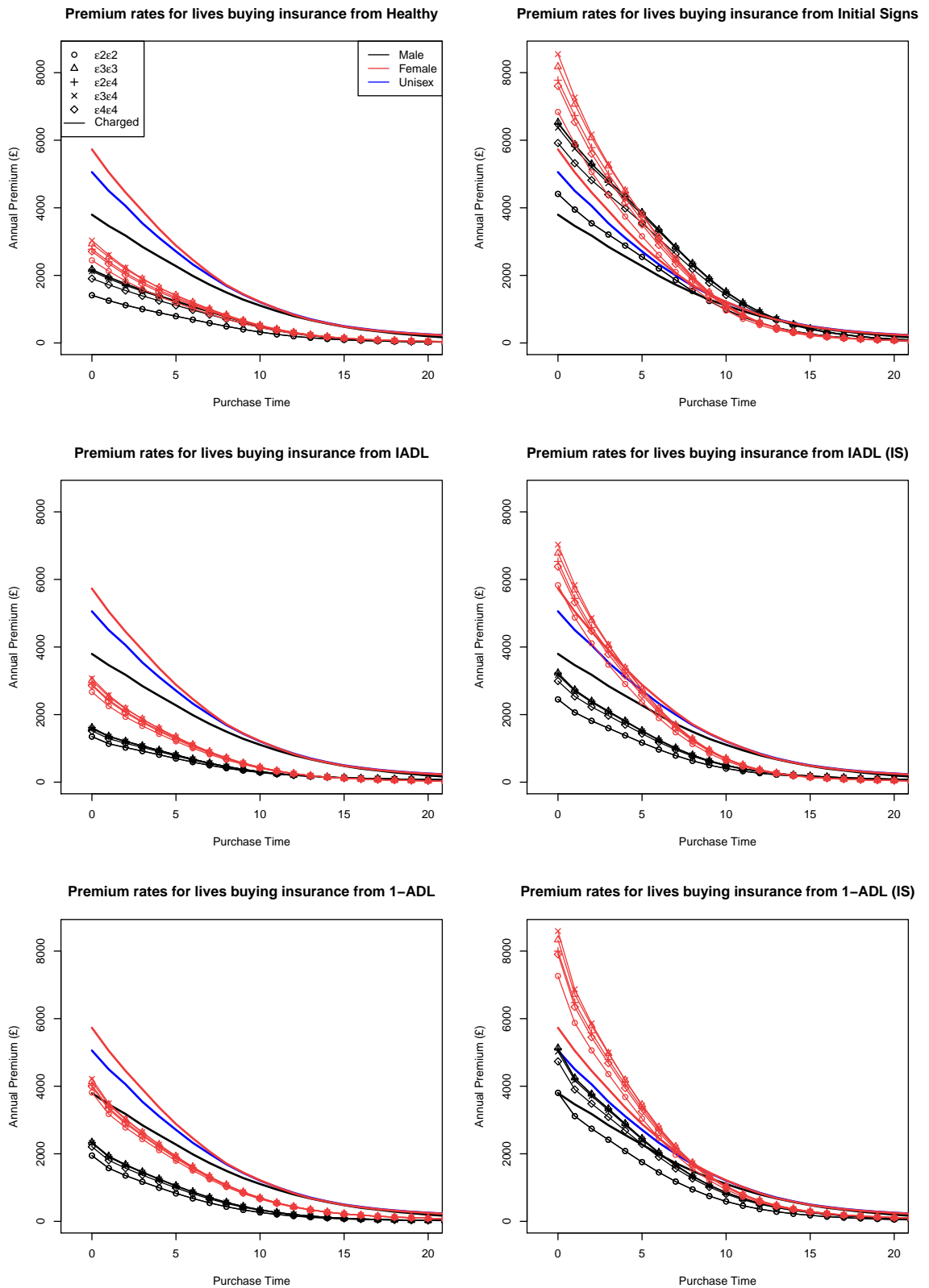
Figure D.6: Single premium rates for LTC, assuming the insurer had complete information about the customer and put lives in each age group, sex, genotype and health status into separate underwriting classes, for lives aged 87.5 at 1st January, 2013. The same legend is used throughout the plots.