THE FINANCIAL IMPACT OF GENETIC INFORMATION ON THE INSURANCE INDUSTRY

By

Fei (Billy) Yu

Submitted for the degree of
Doctor of Philosophy
on completion of research in the
Department of Actuarial Mathematics and Statistics,
School of Mathematical and Computer Sciences,
Heriot-Watt University

February 2010

I hereby declare that the work presented in this thesis was carried out by myself at Heriot-Watt University, Edinburgh, except where due acknowledgement is made, and has not been submitted for any other degree.

Fei Yu (Candidate)

Professor Angus S. Macdonald (First Supervisor)

Dr George Streftaris (Second Supervisor)

Date

Contents

List of Tables

List of Figures

Acknowledgements

I am in great debt to my supervisor Professor Angus Macdonald for his guidance, advice and motivation during the period of my research. His help in finding my internship opportunity at Barrie & Hibbert Ltd was truly invaluable to my future career in Actuarial Prfession. I also want to thank Professor Howard Waters for his encouragement and support during my application for this Ph.D position. In addition, I must note that my being able to apply for ad successfully attaining this Ph.D position would not have been possible without the help and encouragement of Ms Andrea Sneddon during my MSc course in Actuarial Science.

The projects in this thesis were carried out at the Genetics and Insurance Research Centre at Heriot-Watt University, and I would like to thank the sponsors for the funding.

I would also like to thank my wife for her understanding the importance of my work and being here with me in Edinburgh. Parents from both sides of my family gave me great support and encouragement and I also owe them a huge debt for my success.

This thesis aims to accomplish the project: quantification of the overall impact of genetic information on the insurance industry. This project could not have been finishd without a number of other people's contribution and I want to sincerely thank them for their work. They are Delme John Pritchard, Chessman Tavarwisa Wekwete, Eng Hock Gui, Cristina Gutiérrez, Carolina Espinosa Castañeda, Li Lu and Baopeng Lu.

In addition, I would like to thank Dr. Patrice Watson and the colleagues for providing

the data for the study of post-onset mortality associated with colorectal cancer, without which this thesis would not have been possible.

Abstract

This thesis discusses the overall impact of genetic information on the insurance industry using the "bottom-up" approach, in which individual studies of each genetic disorder of interest are studied first. We review five relevant individual studies of adult polycystic kidney disease (APKD), early-onset Alzheimer's disease (EOAD), Huntington's disease (HD), hereditary non-polyposis colorectal cancer (HNPCC), breast & ovarian cancer (BC & OC), and use myotonic dystrophy (MD) to exemplify the methodology of the "bottom-up" approach, and bring these together to quantify the overall impact of genetic information in both the critical illness (CI) insurance and the life insurance market. We also carry out an individual study of OC in the income protect insurance (IPI) market. We conclude that in most cases the cost of adverse selection is negligible and should not cause significant concerns for insurers, especially when we consider other factors, e.g. development of health care and general trend of mortality improvement, which greatly overwhelm the genetic risk. Further this thesis models the colorectal cancer (CRC) screening program as an example and conclude that the CRC screening program appears to reduce the genetic risk by about the same magnitude.

Introduction

We begin by reviewing some facts of human inheritance and genetic testing in Chapter 1. We then look into the issues of genetic information and insurance, and development of these issues in the UK. Some relevant actuarial work and our research objectives are also discussed in Chapter 1. The impact of genetic information on the insurance industry is the central question we address. To answer this question, the "top-down" approach has been completed in Macdonald (1997, 1999 & 2003b) based on high-level general assumptions about the genetic morbidity, with no attempt to model the effect of particular genes. One purpose of this thesis is to quantify the overall impact of genetic information on the insurance industry using the "bottom-up" approach, as opposed to the "top-down" approach. In the "bottom-up" approach, the impact of each genetic disorder is to be studied individually based on the relevant epidemiological studies. By "each genetic disorder", we do not mean all genetic disorders known to us, but only selected major genetic disorders, taken for their significant impact on the insurance industry. The Association of British Insurers (ABI) made a list of eight genetic disorders. Taking the list as the start, with some adjustments, we chose six genetic disorder to study in this thesis, including adult polycystic kidney disease (APKD), early-onset Alzheimer's disease (EOAD), Huntington's disease (HD), hereditary non-polyposis colorectal cancer (HNPCC), breast & ovarian cancer (BC & OC), myotonic dystrophy (MD). Our research work starts with MD and also uses MD to exemplify the methodology of the "bottom-up"

approach.

In Chapter 2 we review some important features of Myotonic Dystrophy (MD). MD is one type of muscular dystrophy, characterized by the slow relaxation of the muscles after contraction, especially hand muscles. MD is inherited in an autosomal dominant manner and caused by an abnormal expansion of a trinucleotide (CTG) repeat. Using the data we collect in relevant papers, we estimate the onset rate, post-onset mortality rate and prevalence rate associated with MD. We then apply these rates in a semi-Markov model of critical illness insurance (CI) and life insurance in order to test the effects of genetic test results or family history on premium ratings. We then extend the model into market models to calculate the cost of adverse selection, measured by the increase of insurance premiums, under various moratoria on the use of genetic information.

In Chapter 3, we review similarly the other five genetic disorders. In Chapter 4, we bring all these together to quantify the overall impact of genetic information of all six genetic disorders, following the methodology introduced in Chapter 2. In Chapter 4, we also discuss the adequacy of the "bottom-up" approach, e.g. the necessity of including more genetic disorders in our list. Chapter 2, 3 and 4 could be regarded as the closure of the "bottom-up" approach.

In Chapter 5, we look at colorectal cancer (CRC) and the CRC screening program. The CRC screening program includes the Bowel Screening Program for the general population and the HNPCC Surveillance Program for HNPCC families. The purpose of a CRC screening program is to detect CRC at early stages so the CRC associated risks can be reduced. We study the effect of a CRC screening program on the onset rate of CRC and the mortality caused by CRC and their subsequent effect on life insurance premium rates.

In Chapters 6 and 7, we look at ovarian cancer (OC) and income protection insurance (IPI). This work is an extension of the study of BC and IPI, Lu, Macdonald & Waters (2008) and Lu *et al.* (2008). In Chapter 6, we introduce some important features of

OC relevant to the pricing of IPI contracts, e.g. diagnosis, treatment, recovery rate and recurrence rate; and we propose a semi-Markov model to model the life history of a woman at risk of OC. In Chapter 7, we use this model to calculate IPI premium rates based on genetic test results or family history, and then extend this model to an IPI market model, so as to estimate the costs of adverse selection under various moratoria on the use of genetic information relevant to OC.

In Chapter 8, conclusions are drawn and further research work is proposed.

Appendix A details the model for critical illness (CI) insurance based on medical studies and population data (Gutièrrez & Macdonald, 2003). Appendix B shows an expanded CI insurance market model of MD. Appendix C has 5 sections, each of which shows the technical details of genetic disorders of interest in this thesis. In Appendix D, we briefly introduce Fourth-Order Runge-Kutta methods.

Chapter 1

Background

1.1 Genetics and Genetic Disorders

1.1.1 Genetics and the Mechanism of Reproduction

Every human has in almost every cell nucleus 23 pairs of chromosomes, 46 in total, including 22 pairs of autosomes, and 1 pair of sex chromosomes. Of each pair, one copy comes from the father and the other copy from the mother, i.e. these two copies contain different information, although they have the same structure. One of the main functions of chromosomes is to pass the paternal and maternal information, carried by sperm and egg respectively, to their child(ren), to this end, each sperm or egg contains only 23 chromosomes. During the production of the sperm and the egg, the two copies of each pair of the chromosomes exchange genetic information, i.e. each of the 23 chromosomes inside the sperm or the egg, is a hybrid of the genetic information carried by the two copies of each pair of the chromosomes in the parent. Note that the exchange of the genetic information increases the variability of the chromosomes.

Each chromosome consists of two strands of nucleotides connected together. Each strand is a sequence of A, T, C, G nucleotides, e.g....-A-T-C-G-C-A-G-A-C-.... The other

strand will be decided by the complementary A-T, C-G link[1]. For example, if one strand is ...-A-T-C-G-C-A-G-A-C-..., the other strand is precisely ...-T-A-G-C-G-T-C-T-G-.... A sequence of triples of nucleotides, e.g. -T-A-G-, determines the sequence of amino acids in the corresponding protein, which is the basic unit of all life. Genes consists of a series of these sequences that encode information for the synthesis of proteins.

A correct sequence of -A-, -T-, -C-, -G- nucleotides leads to the correct production of a protein. When a mutation happens, e.g. -T- in the sequence -A-T-C-G-C-A-G-A-C- is knocked out and replaced by -A-, -C- or -G-, an altered triple may lead to the sequencing of a different amino acid, and consequently the production of a specific faulty protein. Mutations could happen to both copies of a gene, one on each of a pair of chromosomes. The type of faulty protein differentiates two kinds of single gene disorders: recessive disorders and dominant disorders.

For the first type, recessive disorders, the faulty protein produced is harmless. A mutation carrier will not be affected, since one copy fails to produce the correct protein, while the other copy still functions well. Only when both chromosomes carry a mutated gene, will this person develop the genetic disease, and often persons affected have the syndrome at a very early age, since no correct protein is produced since birth.

For the second type, dominant disorders, the faulty protein produced is harmful. The wrong protein produced may be attributable to a mutation in either copy of the gene. Since one copy is still functional, dominant disorders often cause late onset genetic disease. Mutations in both copies of a gene associated with a dominant disorder are rare, and bring about the same effect as a recessive disorder, which means that inheriting two mutations will also lead to early onset disorder; or be incompatible with life.

[1]By A-T, C-G link, we mean that the nucleotide A bonds to T, and the nucleotide C bonds to G, exclusively.

1.1.2 Genetic Disorders

Genetic disorders can also be classified according to the way in which genetic material contributes:

1. Monogenic disorders: These are genetic disorders, often rare, where a specific mutation in a single gene causes a particular condition. Some mutations happen in autosomal chromosomes, causing disorders that may affect both males and females. Some mutations happen in sex chromosomes, resulting in sex-related genetic disorders, which only affect males or females. Monogenic genetic disorders can also be divided into dominant and recessive disorders, which mechanism is described in Section 1.1.1.

2. Multi-gene disorders: These are conditions in which one mutation itself is not enough to trigger the disease, but requires contributions from the environment and also interactions with other genes. These disorders include some common diseases — much heart disease and all cancers.

3. Chromosome disorders: These are conditions caused by the addition or deletion of a whole chromosome or parts of a chromosome. For example, Down's syndrome is caused by a complete extra copy of chromosome 21.

Penetrance means the probability that a mutation carrier would develop the disease by age x, if all other decrements were absent (Macdonald, 2003a). Few genetic diseases have complete, i.e. 100%, penetrance. In the case of incomplete penetrance, the mutation carriers can remain asymptomatic at very old ages.

1.2 Genetics and Insurance

1.2.1 Genetic Information, Genetic Testing and Family History

In a broad sense, as mentioned in ALRC (2003), "almost all information about a person's health and physical well-being can be called 'genetic information'." Adopted from Lu (2006), there are three levels at which information may be regarded as 'genetic information' as follows:

(a) A narrow definition usually refers to DNA-based genetic analysis looking for mutations which are associated with certain diseases.

(b) A broad definition includes family history information about the medical conditions of family members. It may also include the genetic test results of family members.

(c) A sweeping definition may include any inherited characteristics, such as sex, height or hair colour.

In the first annual report by the Genetics and Insurance Committee (GAIC) in 2000, it was said that the term "genetic test" is not easy to define. GAIC adopted the definition used by the Advisory Committee on Genetic Testing, that "a genetic test is a test to detect the presence or absence of, or change in, a particular gene or chromosome." Also in this paper, it said that genetic testing includes biochemical analysis, which detects an inherited change in a gene in an indirect way by detecting a change in the level of a specific gene product (protein), and the analysis of heritable genetic material itself, which is DNA or chromosomes. Several important terms are: predictive genetic testing (which can identify mutations that increase a healthy person's chances of developing disorders with a genetic basis); pre-symptomatic testing (which can determine whether a person is at risk of developing a genetic disorder before any signs or symptoms appear); and in

contrast conventional diagnostic testing (which provides information about the patient's current state).

1.2.2 Underwriting Practices

Several aspects of underwriting practices were reviewed in the first report by GAIC in 2000.

1. The premium for a life (or health) insurance policy comprises three main elements: risk premiums, expenses and profit, in which risk premium is decided by the applicant's age, sex, status of smoking, the term of the policy and other related factors.

2. Two main roles of an underwriter in a life (or health) insurance company are:

   (a) to identify the potential for adverse selection, which arises when insurance applicants have better information than insurance companies, and consequently may buy larger amounts of insurance cover than they would otherwise do.

   (b) to assess extra premium loadings when applicants have extra risks, which exclude them from the standard risk pool.

So the disclosure of genetic testing results might put people in a higher risk group, resulting in higher risk premium loadings. In the case of life insurance, 95% of individual are accepted at a standard rate, about 4% of applicants are accepted at an increased rate to reflect the level of additional risk (typically in the range of +50% to +500%, i.e. 150% to 600% of the standard premium rate), and about 1% of applicants are denied because the risk they carry is exceptionally large. In the case of health insurance, underwriters take similar considerations into account, but with narrower range for acceptable additional premiums (typically, +25% to +200%, i.e. 125% to 300% of standard premium).

1.2.3 Concerns Related to Genetics and Insurance

Concerns about genetics from both insurers and individuals were presented in HGAC (1997). Two important principles, namely solidarity and mutuality, relate to the insurance industry. The best example of the first principle in the UK is the National Health Service and Social Security, in which everybody is treated equally, and is included in the scheme. The best example of the second principle is commercial insurance business. In this case consumers have the option, but no obligation, to purchase the insurance products provided by an insurance company. The insurance company also has the option to deny an insurance applicant on a sensible ground, and has discretion in premium-setting.

Individuals are concerned about the affordability of commercial insurance products. Insurance matters to individuals' lives. It links to home purchase and the protection of dependents. So people could be afraid of taking genetic tests, because they might be treated in a discriminatory way by an insurance company, e.g. cover declined because of the significantly high risk. They might fear that a "genetic under-class" will form because of this. Also they might be concerned about possible disclosure of their genetic information to other specialists, e.g. their GP, who will be obliged to release this information to an insurance company upon request. These concerns could deter people from taking genetic tests and this could delay their receiving treatment in good time. In the case of colon cancer, for example, early detection and an appropriate screening program can greatly reduce the mortality associated with this disease. Insurers argue that such discrimination is groundless due to the competitiveness of the insurance industry. The counter-argument is that the complexity and incomparability of insurance policies would make consumers' concerns meaningful.

A major concern of the insurance industry is the possibility of adverse selection, which arises when customers have better information about their health condition than the insurance company. When genetic information is banned from being used by insurers,

they are in a unfavourable place, compared with the insurance applicants, who could take advantage of the insurance company by purchasing larger amounts of insurance cover than they might normally do. The insurance company will have to make up this loss by raising premiums charged to applicants in the standard risk pool. If the increase is sufficiently high, these people will be driven out of the market, leading to an increasing proportion of applicants carrying mutations, and threatening the sustainability of the insurance industry in extreme cases.

1.2.4 Moratoria in the UK

To prevent genetic discrimination, and protect individuals' genetic information from improper use by interested parties, e.g. insurance companies, a series of events has taken place in the UK, starting in 1995, when the Human Genetics Advisory Commission (HGAC) was appointed by the UK Government to investigate all matters arising from human genetics, including insurance.

In HGAC's 1997 report (HGAC, 1997), it commented that a complete ban on the use of genetic tests in insurance would not be appropriate. Instead, they recommended that insurers should respect a moratorium. They demanded that the moratorium, meaning "a temporary stopping of an activity, especially by official agreement", needed to be of at least 2 years' duration, and subject to review after that, since there had not yet been enough evidence accumulated on this subject. They did not detail all the specifications of a moratorium.

In 1998, in response to the HGAC's report, the UK Government concurred that a permanent ban on the use of genetic tests by insurers would not be appropriate (Department of Trade and Industry, 1998). The Government instead welcomed the Code of Practice on Genetics introduced by Association of British Insurers (ABI), in which the basic rules were:

(a) insurers were not allowed to ask anyone to take a genetic test for insurance purposes;

(b) insurers could not use existing genetic test results except for mortgage-linked life insurance, provided the sum assured exceeded £100,000;

(c) insurers could not use favourable genetic test results to offer better than standard premiums.

The ABI also listed eight genetic disorders regarded as significant for insurance, including early-onset familial Alzheimer's disease (EOAD), familial adenomatous polyposis (FAP), hereditary breast and ovarian cancer (BC & OC), hereditary motor and sensory neuropathy (HMSN), Huntington's disease (HD), Multiple endocrine neoplasia (MEN) and Myotonic dystrophy (MD).

In May 1999, the Human Genetics Commission (HGC) was established by the UK Government. The purpose of setting up this organization was to ensure that the Government could have a complete view on genetics, including cutting-edge knowledge, before they made decisions.

In 2000, the Genetics and Insurance Committee (GAIC) was set up by the Government to determine which genetic tests were suitable for use by the insurance industry, and for what specific types of contract. The first application from the ABI to GAIC was a request to use the results of the genetic test for Huntington's disease in connection with applications for life insurance. This application was approved by GAIC in the same year.

In 2001, the Government signed a agreement with the ABI extending the moratorium to five years, starting from October 2001. UK insurers would continue not to require any genetic test results in respect of applications for life insurance products with sums assured of less than £500,000, critical illness insurance with sums assured of less than £300,000 and for income protection with sum assured up to £300,000. Family history was not included in the moratorium, meaning that insurers remained free to use it.

In early 2005, the ABI extended the moratorium for another five years ending in November 2011. Family history was still not included. Three types of moratoria will be discussed in this thesis:

(a) Moratoria on all genetic test results. In this case, the underwriters are allowed to use family history but not genetic test results. So applicants with a family history will be charged a higher premium than those who do not have a family history, who will be charged the Ordinary Rate (OR).

(b) Moratoria on negative genetic test results only. In this case, underwriters are also allowed to use family history. However, when people with a family history find out that they do not carry a mutation, by taking a genetic test, they may be offered the Ordinary Rate. (This is the form of moratorium in force in the UK.)

(c) Moratoria on genetic test results and family history. In this case, insurers cannot use genetic tests or family history, and everyone will be charged the same level of premiums. (This is the form of moratorium in force in Sweden.)

1.2.5 Studying the Impact of Genetic Information on the Insurance Industry

To study the impact of genetic information on the insurance industry, two approaches have been explored in the Genetics and Insurance Research Centre (GIRC) [2]. The first approach is a "top-down" approach, studied in three papers: Macdonald (1997, 1999 & 2003b). The first two papers studied the entire class of multi-factorial disorders and

[2]GIRC was set up in 1999 in the Department of Actuarial Mathematics and Statistics at Heriot-Watt University, Edinburgh. Please go to its official webpage `http://www.ma.hw.ac.uk/ams/girc/` for more information.

the last considered the entire class of single-gene disorders. Some of the features are summarized as follows.

(a) These papers only look at life insurance. Whether the model proposed in the paper could be extrapolated to other products is unknown.

(b) These works are based on very broad assumptions about mortality, without specific reference to individual diseases. We do not expect this to be accurate, but it does give a very broad sense of the level of adverse selection. However, these studies are only useful when extreme assumptions lead to modest conclusions.

(c) Their advantage is that they need only the broadest guidance from genetic epidemiology. However, when some of the assumptions become unrealistic, the "top-down" approach is inappropriate, and a "bottom-up" approach is needed.

The second, "bottom-up", approach starts with studies of individual genetic disorders (for brevity, we call this type of study an individual study), and proceeds by aggregating all these individual studies. In individual studies, epidemiological features of each disease are carefully reviewed, e.g. prevalence rate, onset rate and post-onset mortality rate. Markov or semi-Markov models are used to model various insurance products, incorporating the epidemiology of the disease. Using these, we can assess the financial impact of genetic information relating to the disease on insurance applicants, and insurance companies. The financial impact is two-fold based on the concerns we discussed in Section 1.2.3. From an applicant's perspective, we calculate the premium rates for mutation-carriers, expressed as a percentage of that for non-mutation carriers, and then the insurability of insurance applicants with mutations is assessed based on underwriting practices we introduced in Section 1.2.2. We can extend this analysis to applicants who have a family history of the disease but whose genotype is unknown. From an insurer's perspective, we further calculate the possible cost of adverse selection. Although the "bottom-up" approach

requires the studies of each individual genetic disorder, it is clearly impossible to include all known genetic disorders. As mentioned in Section 1.2.4, the ABI listed eight genetic disorders regarded as significant for insurance, including early-onset familial Alzheimer's disease (EOAD), familial adenomatous polyposis (FAP), hereditary breast and ovarian cancer (BC & OC), hereditary motor and sensory neuropathy (HMSN), Huntington's disease (HD), Multiple endocrine neoplasia (MEN) and Myotonic dystrophy (MD). Not all of these disorders have been made the subject of mathematical modelling in GIRC, for various reasons as follows:

(a) MEN is an example of a cancer for which genetic testing and early treatment should lead to substantially better outcomes (Gui, 2003), hence we may doubt that it would have a substantial impact on the insurance industry. Please see Gui (2003) for more details about this disease.

(b) HMSN is a very rare disease and most patients affected are very young, usually below age 20, before they would be likely to purchase insurance. Therefore this disease is proven not to have a substantial impact on the insurance industry, and so has not been modelled.

(c) FAP is a type of hereditary colorectal cancer, which has a feature of polyps. Because these polyps can be easily detected and removed in a screening program, the risk associated with FAP is limited, and like MEN, ought to be substantially lowered by genetic testing within at-risk families. Therefore, FAP is not included.

Some diseases not on the ABI list, however, have been modelled in GIRC as follows:

(a) Hereditary nonpolyposis colorectal cancer (HNPCC) is fully studied in Lu *et al.* (2007).

(b) Adult polycystic kidney disease (APKD), originally on the ABI list, was later removed because it is typically detected by ultrasonography, which is not DNA-based

genetic testing according to the narrow definition of genetic testing the ABI adopted. But because APKD is one of the most common dominant single-gene hereditary diseases, APKD is studied in Gutiérrez & Macdonald (2003 and 2007).

After these adjustments, six genetic diseases, including adult polycystic kidney disease (APKD), early-onset Alzheimer's disease (EOAD), Huntington's disease (HD), myotonic dystrophy (MD), hereditary non-polyposis colorectal cancer (HNPCC) and breast cancer & ovarian cancer (BC & OC), were chosen to be studied in GIRC. A series of papers, i.e. individual studies in the 'bottom-up' program, was published as follows:

(a) APKD: Gutiérrez & Macdonald (2003) and Gutiérrez & Macdonald (2007).

(b) EOAD: Gui & Macdonald (2002) and Espinosa & Macdonald (2007).

(c) HD: Gutiérrez & Macdonald (2004).

(d) MD: This work will be presented as Chapter 2 of this thesis.

(e) HNPCC: Lu *et al.* (2007).

(f) BC & OC: Macdonald, Waters & Wekwete (2003a & b) and Gui *et al.* (2006).

The conclusions made in individual studies are as follows.

(a) The carriers of specific mutations, healthy at the time of applying for insurance, are not insurable, with few exceptions.

(b) Individual genetic disorders will not result in substantial adverse selection costs for the insurance industry.

We observe from the above list that no individual study of MD has yet been done. So Chapter 2 will carry this out. This done, we are ready to model the overall cost of adverse selection (for brevity, we call this part the aggregate study).

Chapter 2

Myotonic Dystrophy

2.1 Introduction to Myotonic Dystrophy

2.1.1 Muscular Dystrophy

Muscular Dystrophy represents a group of muscular disorders characterized by progressive muscle weakness and loss of muscle tissue. Over 30 muscle diseases fall into this category. Of these, 9 types are well known and named specifically. They are: Becker's muscular dystrphy (BMD), Duchenne muscular dystrophy (DMD), Emery-Dreifuss muscular dystrophy (EDMD), Limb-girdle muscular dystrophy (LGMD), Facioscapulohumeral muscular dystrophy (FSHMD), Myotonic dystrophy (MD), Oculopharyngeal muscular dystrophy (OPMD), Distal muscular dystrophy (DD), and Congenital muscular dystrophy (CMD). The first 3, BMD, DMD and EDMD are sex-related, only affecting males, because the mutations are in genes on the X chromosome. Females are carriers, but rarely affected. The other 6 could affect both males and females because the mutations occur in genes on autosomal chromosomes. Of these 9 well-defined types of muscular dystrophy, the severity and the onset age vary largely. The earliest onset case is CMD. Childhood and early adulthood onset cases include BMD, DMD, EDMD and FSHMD. LGMD patients

could be affected from late childhood to middle age. Adult onset cases are DD, MD and OPMD. MD is the most common form of adult onset muscular dystrophy.

2.1.2 A Brief History of Myotonic Dystrophy

Myotonia, characterized by the slow relaxation of the muscles after contraction, especially hand muscles, was first described as such by Dr Julius Thomsen in 1876. MD was first described by a German physician, Dr. Hans Steinert in 1904 (Harper, 2001). He studied 9 cases, noting symptoms such as muscle weakness and wasting, facial weakness, etc. Dr. Hans Steinert first used the phrase "myotonic atrophica" to describe this disorder. In 1909, shortly after the premature death of Dr Hans Steinert, Batten and Gibbs pursued the work. They expanded the literature on MD by publishing 5 cases and reviewing the published cases of myotonic dystrophy and realized that all these cases could form a definite disorder group. In 1911, Greenfield was the first person who recognized cataract to be another form of onset of MD, after studying a sibship of 13 with cataracts, and tracing cataracts in the paternal aunt and grandmother. This work can be recognized as the first phase of the study of MD, including its discovery, establishment of the concept, and the initial collection of clinical documents. The second phase of this work focused more on the inheritance pattern, which the clinical aspects were expanded. The first systematic study of MD was carried out by Bell (1947) and Thomasen (1948). In their work, they established autosomal dominant inheritance. The third phase of study concentrated more on the molecular basis of MD. The first mapping of gene locus took place in 1971, and in 1992, the cause of myotonic dystrophy was finally discovered to be an expanded trinucleotide (CTG) repeat sequence.

2.1.3 The Genetic Mechanism and Classification of Myotonic Dystrophy

As we said in the previous section, MD is caused by an abnormal expansion of a trinucleotide CTG repeat. There are three forms of MD: type 1 (MD1), type 2 (MD2) and type 3 (MD3). In MD1, the expansion is in a non-coding region of the DMPK (Dystrophia-myotonica protein kinase) gene on chromosome 19. In MD2, the expansion is in a non-coding region of the ZNF9 (zinc finger protein 9) gene on chromosome 3. Very few papers refer to the rare MD3, and we will not consider it here. MD1 is the most common form, accounting for about 98 percent of all cases. MD2 has milder clinical presentation than MD1. MD3 is newly found and even rarer than MD2. Therefore, here we will only consider MD1.

2.1.4 Classification of Myotonic Dystrophy 1

There are 3 forms of MD1: congenital MD (CMD); classical MD; and mild MD. These forms are differentiated by the number of CTG repeats. Normal people have fewer than 50 CTG repeats. Mild MD1 is usually associated with 50–100 CTG repeats, classical adult onset with 100–1,000 CTG repeats and the more severe CMD with over 1,000 CTG repeats. (The classification varies in different parts of the literature.) Table 2.1 shows a typical classification (Langlois, 2003).

Similar tables can be found elsewhere, often set up for different purposes. A general pattern is that the bigger the number of CTG repeats, the earlier the onset will be. CMD is more severe than other two cases, and the onset at birth makes the carriers uninsurable. Therefore in the following paper, we concentrate on the adult onset cases of MD1.

Table 2.1: Critical CTG expansion thresholds for myotonic dystrophy onset and symptoms. Source: Langlois (2003).

No. of CTG repeats	Phenotype	Comments
3–49	Normal Individual	Over 90% of the general population has fewer than 35 CTG repeats at the MD1 locus.
50–80	Very mild or no apparent symptoms except cataracts	Expansions in this range are incremental from generation to generation.
80–1000	Full mutation, adult onset	80 CTG repeats is the threshold for saltatory amplification.
≥ 600	Full mutation, adult onset	High risk of transmitting the congenital form if the mother is the carrier.
1000–3000	Generally congenital	Such expansions are most frequently seen in congenital cases.

2.1.5 Clinical Aspects of Myotonic Dystrophy 1

One significant feature of the clinical aspects of MD 1 is the occurrence of a variety of abnormalities outside muscle tissue. Höweler (1986) gave a diagram, describing the severity of three typical symptoms: cataract, muscle weakness and mental symptoms, categorized by different age of onset. Two features were that mental retardation mostly happens in CMD and young adulthood, while cataract mostly affects people in the late onset group. Mild MD normally comes with cataract and light muscle stiffness. Facial muscles are one

of the frequently affected muscle groups, weakness in which could cause motionlessness of patients. Ptosis, the medical term for drooping eyelid, is another common symptom; affected people might not be able to open their eyes. Other symptoms, e.g. diplopia, more commonly known as double vision, are directly attributable to extraocular muscle weakness. Jaw and tongue muscles are also commonly affected by MD1. All these facial muscle weaknesses, along with mental retardation, very often cause speech problems in CMD cases. Weakness of limb muscles, especially distal muscles [1] of the forearm is usually associated with MD1. Myotonia, as a hallmark of MD, refers to the fact that affected patients find it very hard to relax their hand after grasping an object. Extra-muscular abnormalities involve other inner organs of the human body, e.g. heart, lungs, endocrine system, brain, and skin.

The most comprehensive assessment of physical disability associated with MD1 is Mathieu *et al.* (1992). In this paper, the authors used the muscular disability rating scale (MDRS) as displayed in Table 2.2, to evaluate 295 MD patients living in the Saguenay-lac-Saint-Jean (SLSJ) region (Quebec, Canada), which has the highest reported prevalence of MD in the world, 189 per 100,000 population.

Table 2.2: Muscular disability rating scale (MDRS).

Grade	Description
1	No clinical muscular impairment
2	Minimal signs (myotonia, facial weakness and etc)
3	Distal weakness
4	Mild or moderate proximal weakness
5	Severe proximal weakness (confined to wheelchair)

[1] The muscles further from the centre of the body are called distal muscles and the muscles closer to the centre of the body are called proximal muscles

This conclusion on the duration of the disease is that on average, a distal weakness (grade 3) is identified after 9.6 years' duration (SD = 5.9 years), a proximal weakness (grade 4) is noticed after a progression of 18 years (SD = 9.1 years), and patients are wheelchair-bound (grade 5) after a course of 27 years (SD = 9.9 years). The authors noticed significant correlation between the duration of the disease and the severity.

2.1.6 Age at Onset

Mathieu *et al.* (1999), another study of MD patients in the SLSJ area, found that the mean age at onset for males is 17.0, and for females is 20.0. However, these means are all subject to high variances. The range for males is 0–44, and that for females is 0–51. In Lynas (1957), the mean of age at onset is 31.1, and the range of age at onset is 7–70. Other papers, e.g. Maas (1937) and Bell (1947) show similar conclusions.

2.1.7 Negative Correlation between Age at Onset and Number of CTG Repeats

Several reports agreed on a broad conclusion that the age at onset has a negative relationship with the number of CTG repeats (Hunter *et al.*, 1992; Novelli *et al.*, 1993). This phenomenon coincides with the clinical aspects of MD we discussed previously. Hamshere *et al.* (1999) recently found that the correlation of (CTG) repeat length in leucocytes, with age at onset, is significant only for patients with small expansions. In Hsiao *et al.* (2003), 96 subjects belonging to 26 families were identified as MD1 patients. The inverse correlation between the age at onset and the number of the CTG repeats only existed for small expansions (≤ 250). This opinion is very similar to Hamshere *et al.* (1999).

2.1.8 Age at Death and Cause of Death

Thomasen (1948) found the mean age at death of 24 patients to be 43.5 years. Bell (1947) found the mean age at death of 85 published cases to be almost the same at 44.7 years, compared with approximately 60 years for the population as a whole over the same period. Klein (1958) recorded a somewhat higher mean age at death (50.6 years), which may reflect the generally improved life expectancy since the publication of Bell (1947). This result was in agreement with Grimm (1975a).

Two papers have contributed to the measurement of MD-associated death, de Die-Smulders *et al.* (1998) and Mathieu *et al.* (1999). In the first paper, the causes of MD-related death were sudden death, arrhythmias, pneumonia, and postoperative death. In the second paper, of 75 patients who died during the observation, 32 died from a respiratory problem, either pneumonia or acute respiratory failure, and 15 patients from a cardiovascular disease. The latter result agrees with the former result.

2.1.9 Prognosis

Estimating the age at onset, the age at death and the cause of death, and in general the prognosis for patients with MD are not easy tasks because the age at onset, the onset symptoms, the progression of the disease, the weakness of muscles, the age at death and the cause of death vary dramatically (Harper, 2001).

2.1.10 Diagnosis of Myotonic Dystrophy 1

Accurate diagnoses of MD1 and MD2 are most likely to be made by doctors with specialized training in both neurology and the adult neuromuscular system. Several factors might explain the difficulty of accurately diagnosing MD.

(a) In most cases, patients are treated by family doctors, who might not be aware that

these symptoms are related to MD1 or MD2.

(b) Depending on the presentation of symptoms, people may be referred to other medical specialists including cardiologists, endocrinologists etc. However, they might not be aware of the criteria or the full range of clinical symptoms of MD, which is a multi-system disease. These specialists might give opinions based on their experience but fail to make a collaborative diagnosis.

(c) Neuromuscular disorders covers more than 40 different diseases. All these diseases might share some common characteristics, e.g. progressive degeneration of muscle systems. Before some textbook symptoms of MD appear, even neurologists find it hard to tell what disease it might be.

(d) Not all delayed relaxation of muscle is the result of myotonia, and a number of conditions may cause confusion. Other diseases, e.g. other muscular dystrophies introduced previously, can also cause great confusion.

Therefore, clinical presentation is not now enough for an accurate diagnosis without the help of a molecular test. However, that was not the case before molecular analysis was available. Harper (2001) introduced two clinical tests offering presymptomatic detection: namely the early detection of lens opacities by slit-lamp examination, and the recognition of subclinical myotonia by electromyography. Both are now recognized to have significant false-positive and false-negative rates. They may still be useful if a confirmed mutation carrier is being monitored regularly and wishes to know if there is any evidence of early ocular or muscle disease.

Since the emergence of the molecular test, the genetic test has become the most convincing and reliable result in diagnosis of MD. In Brunner *et al.* (1992), the authors corrected the diagnoses for 11 subjects who had been ascertained by genetic linkage studies, and found that 9 subjects had expanded CTG repeats and 2 subjects had normal

numbers of CTG repeats. Shelbourne *et al.* (1993) concluded that the genetic test provided better diagnoses. The genetic test method used in these two papers were Southern blot analysis and polymerase chain reaction (PCR), respectively.

These two methods have their own advantages and disadvantages. The Southern blot method is used to detect larger than 100 CTG repeats. So this is a standard procedure to detect MD1 mutations. But when the number of CTG repeats is below 100, it must be identified by PCR, which cannot provide reliable results when the region of CTG repeats is longer than 500 base pairs. So PCR is not suitable for the detection of severe MD1 mutations (Ashizawa, 2000).

2.1.11 Treatment for Myotonic Dystrophy

Unfortunately, there is no known cure for any type of muscular dystrophy (Harper, 2001). Inactivity might worsen the disease. Physical therapy and orthopedic instruments (e.g. wheelchairs, standing frames) may be helpful. But the aim of physical therapy and drug therapy is not to prevent disease progression, but to alleviate the suffering associated with it.

In Hashem & Sinden (2002), the authors talk about the treatment for the diseases caused by abnormal expansion of DNA repeats, including Huntington's disease (HD), MD etc. They propose to introduce damage into the DNA, which would subsequently lead to an increased rate of repeat deletion. This is very encouraging news for MD patients. However, we are still a long way from any applicable, affordable and easy-to-use therapy.

2.1.12 Anticipation

Anticipation, meaning that each successive generation is normally more severely affected than the last, with earlier onset, is not only associated with MD, but also other hereditary diseases caused by abnormal DNA repeat expansions. Huntington's disease is another

good example. In the case of MD, the number of CTG repeats becomes very unstable in the transmission of DNA through generations. Anticipation was first described by Fleischer in 1918 (Höweler, *et al.*, 1989). Höweler (1986) observed that the children in 60 of 61 parent/child pairs had onset age one or two decades earlier than that of their parents. This again showed the existence of anticipation. However, it is very hard to explain the mechanism of anticipation without the advances in molecular studies of the last decades of the 20th century. Harley *et al.* (1993) and Buxton *et al.* (1992), both found that a variable DNA insert of as much as 5 kb in length, tentatively explained the phenomenon. Also Ashizawa *et al.* (1992) studied CTG expansions in 43 parent/child pairs and observed generally that there is an increasing amplification of the CTG repeats over generations.

Shelbourne *et al.* (1993) published the first report of a transmission from father to son, where the CTG repeat expansion reduced to a size within the normal range, and the son remained asymptomatic. Similar observations are made in Abeliovich *et al.* (1993), Hunter *et al.* (1993), and O'Hoy *et al.* (1993), although this does not happen as frequently as expansion does. The largest study is Ashizawa *et al.* (1994). 1,489 parent/child pairs were studied: 10% of paternal transmissions and 3% of maternal transmissions showed a contraction.

Therefore, in the transmission of DNA, the CTG expansion can be unstable. Most likely, there will be an expansion of CTG repeats. However, there is some probability that a reduction of CTG repeats could happen.

2.2 The Epidemiology of Myotonic Dystrophy

2.2.1 The Penetrance Rate of Myotonic Dystrophy 1

The age-related cumulative risk (penetrance), denoted $F(x)$, can be regarded as the probability of developing MD by age x, if we assume there were no other causes of decrement (Macdonald, 2003a). When x approaches a suitable limiting age, e.g. 100 year old, we obtain the lifetime penetrance. Nesterov *et al.* (1983) claimed to have shown that incomplete penetrance of the MD gene, with penetrance of 83% in Ukrainian families, and 91% in Russian families. Höweler (1986) studied 14 MD families, and found that 46% of subjects were affected. Thus the penetrance of the MD gene was almost complete. Harper (2001) mentioned an earlier study (Harper, 1973) where the ascertainment was through the parent, not the child. Almost 50% proportion of offspring were affected. This implies that MD has almost complete penetrance. Therefore, we conclude that the penetrance of a MD mutation carrier is close to 100%.

2.2.2 The Onset Rate of Myotonic Dystrophy 1

Given estimates of the penetrance $F(x)$, the rate of onset (hazard rate), $\mu(x)$, is then given by:

$$\mu(x) = \frac{F'(x)}{1 - F(x)}, \tag{2.1}$$

which can be computed by analytical or numerical differentiation.

We searched exhaustively the literature on the penetrance of MD. We found several papers carrying detailed pedigrees and descriptions. These include Maas (1937), Bell (1947), Thomasen (1948), Lynas (1957), Harper (1972), Grimm (1975a), Höweler (1986) and Magee (1996). However, these works did not employ survival analysis. So the patients'

information was collected mainly for clinical use, i.e. prognosis and diagnosis, but not for survival analysis. Other authors carried out quasi-survival analysis, calculating the cumulative proportion of the patients affected by MD based on retrospectively collected pedigrees. The pedigrees contained either incomplete, or redundant information from an actuary's point of view.

Frohock (2003) studied MD patients and obtained Kaplan-Meier estimates of the cumulative onset probability. Unlike Huntington's disease, in which there is a relatively limited number of mutation genotypes, the number of CTG repeats in MD could range from 50 to several thousands. So we could not differentiate the genotypes of patients by the exact number of the CTG repeat expansion. In Frohock (2003), the MD mutation carriers are categorized into two groups: genotype I, with the number of the CTG repeats more than 250 (for brevity, this genotype is denoted as CTG250+), and genotype II, with the number of the CTG repeats less than 250 (for brevity, this genotype is denoted as CTG250–). Kaplan-Meier estimates of cumulative onset probability of both genotypes are provided in Frohock (2003). We smoothed these Kaplan-Meier estimates as shown in equations (2.2) and (2.3). Figure 2.1 shows the observed and fitted cumulative onset probabilities, and the corresponding onset rates $\mu_x^{CTG250+}$ and $\mu_x^{CTG250--}$.

$$F(x)^{MD,CTG250+} = \frac{\exp\left(-3.952\times10^{-6}x^3 - 1.624\times10^{-4}x^2 + 0.1206x - 4.951\right)}{1+\exp\left(-3.952\times10^{-6}x^3 - 1.624\times10^{-4}x^2 + 0.1206x - 4.951\right)}, \tag{2.2}$$

$$F(x)^{MD,CTG250--} = \frac{\exp\left(4.343\times10^{-5}x^3 - 0.006044x^2 + 0.4437x - 8.731\right)}{1+\exp\left(4.343\times10^{-5}x^3 - 0.006044x^2 + 0.4437x - 8.731\right)}. \tag{2.3}$$

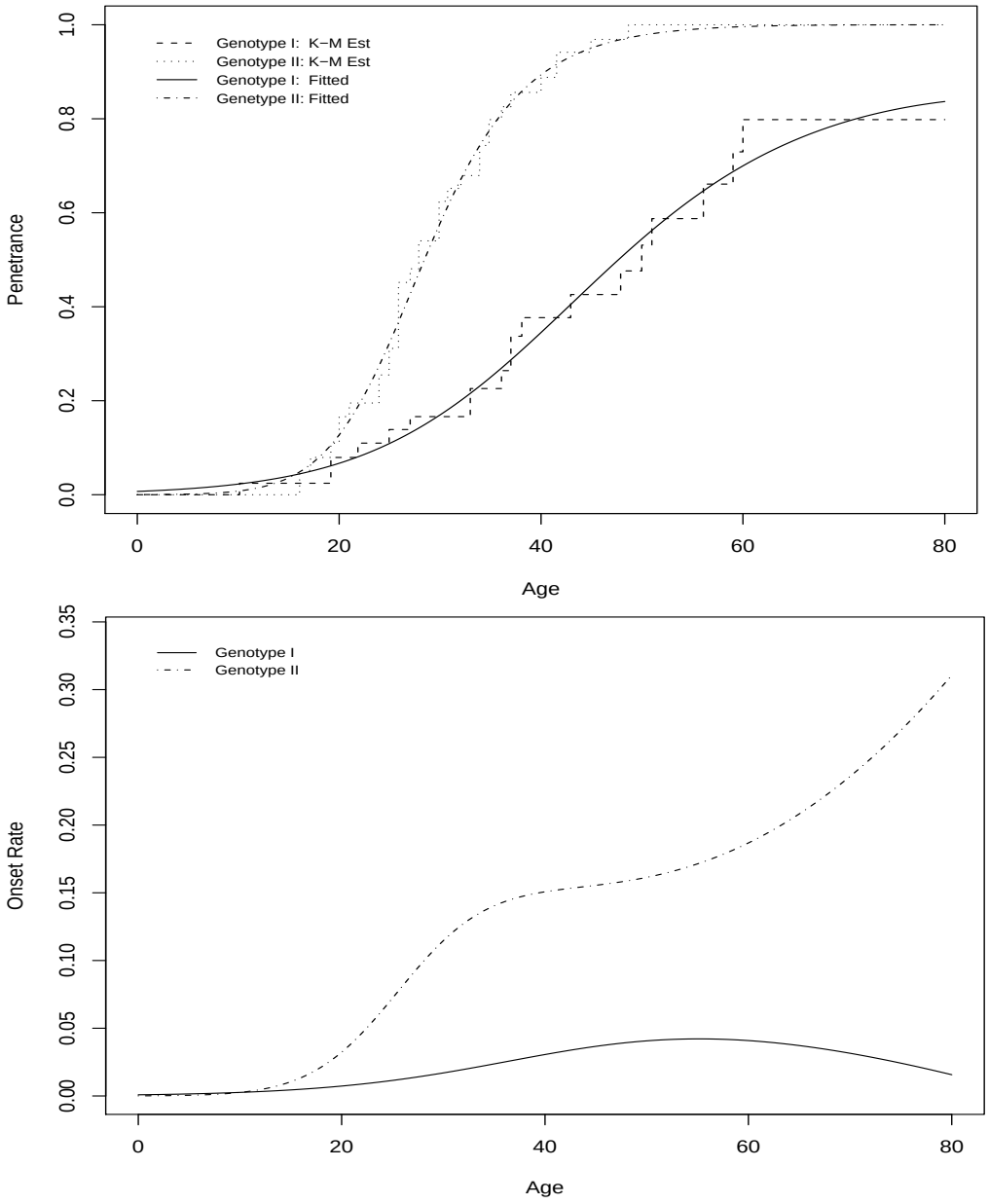

Figure 2.1: Observed and the fitted cumulative onset probabilities (top) and onset rates (bottom) for MD mutation CTG250+ and CTG250– carriers. Source: Frohock (2003) and our fitted curves.

For mutation CTG250+ carriers, since data after 50 years old is scarce and the penetrance is nearly complete by age 50, we can expect a little increase of the onset rate after age 50. Therefore we assume that the onset rate levels off after age 50.

2.2.3 Post-Onset Mortality

On the topic of post-onset mortality, we located three papers, de Die-Smulders *et al.* (1998), Mathieu *et al.* (1999) and Mladenovic *et al.* (2006).

In de Die-Smulders *et al.* (1998), 180 MD patients (from a register of MD, set up in Southern Limburg, Netherlands) were collected for the study of survival after onset. Kaplan-Meier methods were used to calculate the empirical survival probability for both males and females. However, this experiment is not based on the length of follow-up since onset, but the age of patients. This means that at the start of observation, some patients have been ill for some time, but exactly what time is unknown to us. So for this reason, we did not use this paper.

Mathieu *et al.* (1999) and Mladenovic *et al.* (2006) studied data from the Saguenay-Lac-Saint-Jean (SLSJ) region (Quebec, Canada), and the Belgrade area, respectively. As mentioned before, the SLSJ region has the highest prevalence rate of MD in the world. These authors carried out Kaplan-Meier estimation which would suit the needs of our research. However, the drawback of Mathieu *et al.* (1999) is that the author specified four cohorts, Mild, Adult, Early adult and Childhood, but not age ranges for each cohort. Also, the follow-up period in Mathieu *et al.* (1999) is 10 years after onset, which is shorter than Mladenovic *et al.* (2006). Therefore, we adopted the results from Mladenovic *et al.* (2006) because they specify the cohort age ranges, have a longer follow-up period and are more recently published.

Mladenovic *et al.* (2006) provides Kaplan-Meier estimates of the post-onset 15-year duration-dependent survival probability for three cohorts, CMD (onset before age 20),

classical onset (onset between age 20 and age 50), and mild onset (onset after age 50). We are not interested in CMD, because we assume that critical illness insurance and life insurance operate between ages 20 and 60. We decided not to use the mild onset case, because the fitted curve behaves very irregularly due to lack of data and the survival probability by duration 16 years is about the same for both cases. Instead, for both mild onset and classical onset, we use the estimate for classical onset.

We use a logit transformation to fit a smoothed curve to the Kaplan-Meier estimate for the classical onset case. Equation (2.4) gives the fitted function $F(z)^{MD,Mortality}$, as a function of duration z since onset. The post-onset mortality rate $\mu_z^{MD,Mortality}$ is given by equation (2.1). However, Wilkie (2000) pointed out the anomaly that $\mu_z^{MD,Mortality}$ could be assumed to fall substantially below the normal age-related mortality rates $\mu_x^{Standard}$, e.g. English Life Tables 15, at certain ages. Therefore, to avoid this anomaly we assume that mortality after onset of MD is no better than normal age-related population mortality, e.g. $\mu_{x,z}^{MD,Dead} = \max\left\{\mu_z^{MD,Mortality}, \mu_x^{Standard}\right\}$. Using a standard mortality table to adjust the duration-dependent mortality is a technique we will use often in the rest of this thesis. We call this treatment the Wilkie adjustment to mortality. Figure (2.2.3) shows the Kaplan-Meier estimate and the fitted curve of the post-onset mortality probability (top) and the corresponding post-onset mortality rate $\mu_z^{MD,Mortality}$ (bottom).

$$F(z)^{MD,Mortality} = \frac{\exp\left(0.001903z^3 - 0.06907z^2 + 1.007z - 6.082\right)}{1 + \exp\left(0.001903z^3 - 0.06907z^2 + 1.007z - 6.082\right)}. \tag{2.4}$$

According to Equation (2.1) and (2.4), we derive that:

$$\begin{aligned} \mu_z^{MD,Mortality} &= \frac{\exp\left(0.001903z^3 - 0.06907z^2 + 1.007z - 6.082\right)}{1 + \exp\left(0.001903z^3 - 0.06907z^2 + 1.007z - 6.082\right)} \\ &\quad \times (0.005709z^2 - 0.1381z + 1.007) \end{aligned} \tag{2.5}$$

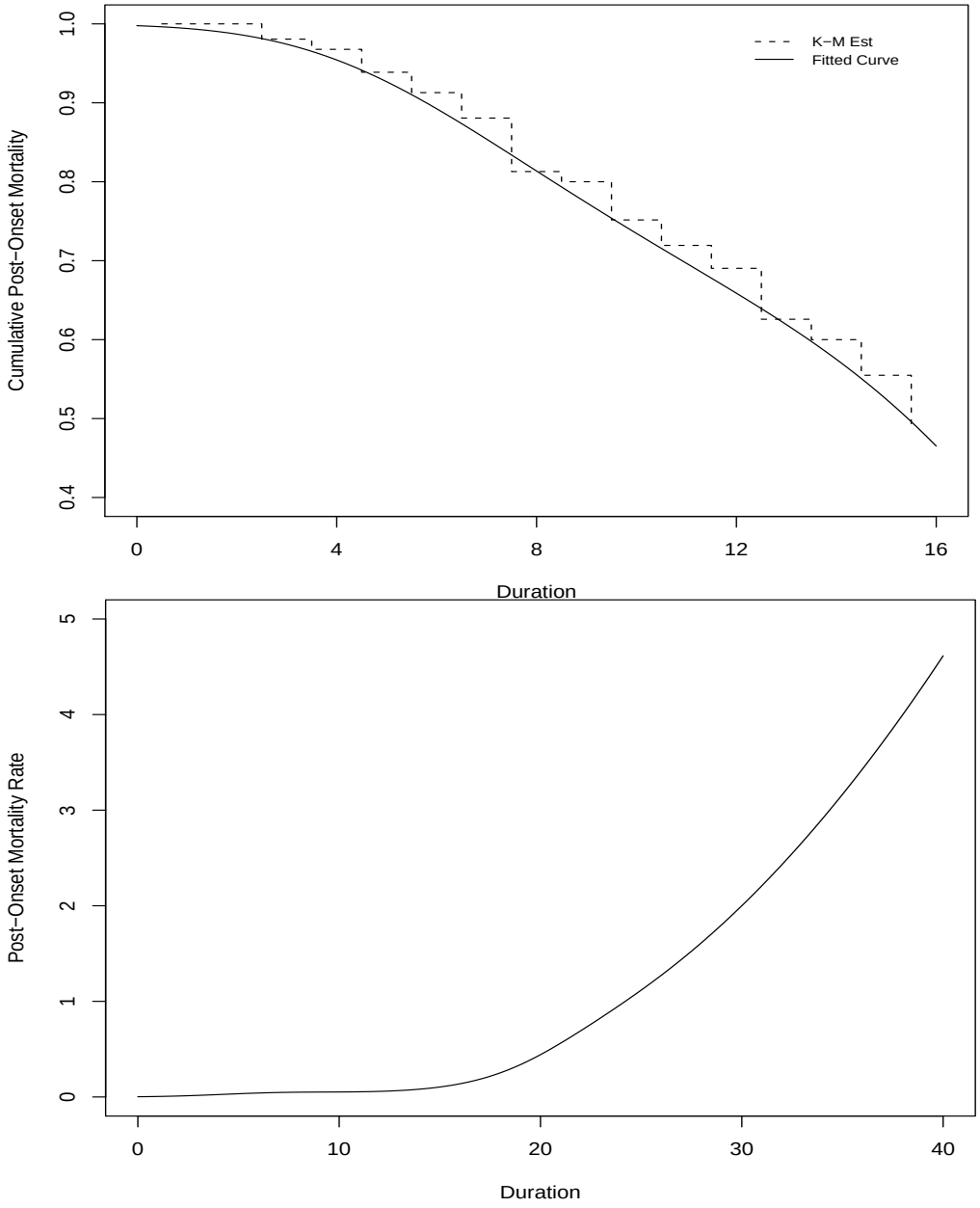

Figure 2.2: Observed and fitted post-onset survival probability $F(z)^{MD,Mortality}$ as a function of the duration of the disease (top) and the post-onset mortality rate $\mu_z^{MD,Mortality}$ as a function of duration since onset of MD . Data source: Mladenovic *et al.* (2006).

2.2.4 Prevalence Rate and Distribution of Mutations

Overall, the prevalence rate of MD is about 1 per 8,000. But prevalence rates in different areas vary hugely. Table 2.3 is a short summary of the prevalence rates in different areas.

Table 2.3: Prevalence of myotonic dystrophy, per 100,000 of population

Source	Place	Frequency
Osame & Furusho (1983)	Kagoshima and Koinawa districts, Japan	5.5
Hsiao *et al.* (2003)	Taiwan	0.46
Ford, Kidd & Hammond-Tooke (2006)	Otago, New Zealand	11.6
Bouchard *et al.* (1988)	Saguenay, Quebec (Canada)	210.5
Grimm (1975b)	Germany	5.5
Mladenovic *et al.* (2005)	central Serbia	3.8
López De Munain *et al.* (1993)	Basque Country, Spain	26.5
Medica, Markovi & Peterlin (1997)	Istria, Croatia	18.1
Klein (1958)	Switzerland	4.9
Mostacciulol *et al.* (1987)	Veneto, Italy	36.3
Siciliano *et al.* (2001)	Padova and North-West Tuscany, Italy	9.31
Magee & Nevin (1999)	Northern Ireland	119.5
MacMillan & Harper (1991)	Wales (South)	7.1

Sizhong *et al.* (2000) reports a very low frequency of MD in Chinese Han people. Mor-Cohen *et al.* (1997) mentioned similar result in Ashkenazic Jews. Goldman, Ramsay & Jenkins (1994) used genetic tests to explain the absence of MD in southern African Negroids. At the other extreme, the Saguenay-Lac-Saint-Jean (SLSJ) region (Quebec, Canada) has 30 to 60 times the worldwide prevalence as shown in Table 2.3. Mathieu,

Braekeleer & Prévost (1990) identified 746 patients distributed in 88 families. They traced all the patients back to a couple who settled in "Nouvelle-France" in 1657. Because of the isolation of this area, the MD gene passed on over 10–14 generations. Since in this paper, we are mainly interested in the UK, we use the estimate of 7.1 per 100,000 (MacMillan & Harper, 1991). This result is also cited in Harper (2001).

The distribution of mutation genotypes is estimated by referring to the data specified in Frohock (2003), which provides the number of the CTG repeat of a sample of patients, including symptomatic patients and asymptomatic patients. We found that 39 patients have genotype CTG250–, and 44 patients have genotype CTG250+. Therefore we assume that the patients have roughly equal chance of having genotype CTG250– and genotype CTG250+ at birth.

2.3 A Critical Illness Insurance Model

2.3.1 Models for Critical Illness Insurance

Since there is no standard critical illness insurance model used by the industry, we adopt the critical illness insurance model from Gutiérrez & Macdonald (2003) as presented in Figure 2.3. This is a continuous-time discrete-state Markov model. In this model, an insured person pays premiums while in State 0, and receives a lump-sum benefit upon transition into State 1 or State 2. This model was used in the case of APKD (Gutiérrez & Macdonald, 2003 and 2007) and EOAD (Gui & Macdonald, 2002 and Espinosa & Macdonald, 2007) for the reason that the onset of APKD or EOAD will trigger the payment for these diseases immediately.

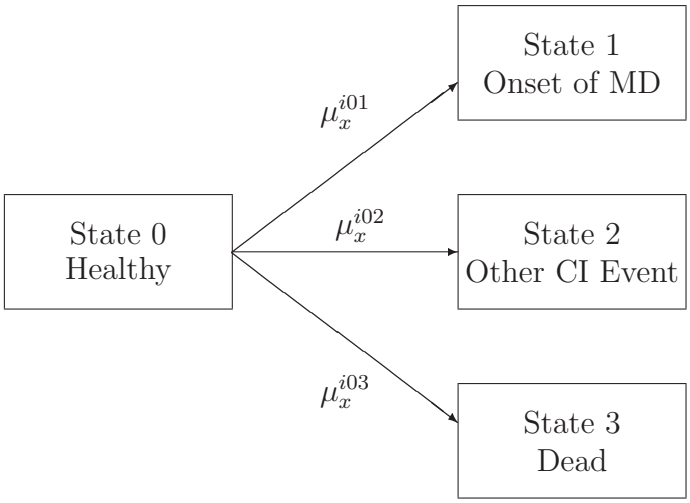

Figure 2.3: A multiple-state Markov model for myotonic dystrophy and in critical illness insurance for an insurance applicant with genotype g_i.

In this model, three intensities as shown need to be parameterized. Intensity μ_x^{i01} is found from the onset rates of MD in Section 2.2.2. Intensities μ_x^{i02} and μ_x^{i03} have been estimated in Gutiérrez & Macdonald (2003). Please see the Appendix A for details.

2.3.2 The Timing of the Critical Illness Insurance Payment

The Extended Model for Critical Illness Insurance

The onset of the MD is associated with less severe symptoms, e.g. myotonia and cataract. These will not trigger the CI insurance payment. In practice, the payment for MD is delayed until some time after onset. This feature could be reflected by the addition of a new state, i.e. State 4: Payment for MD, to the model shown in Figure 2.3. The intensity μ_z^{i14} is the transition rate from State 1 to State 4, in which z is the duration since onset of MD. The new model is shown in Figure 2.4.

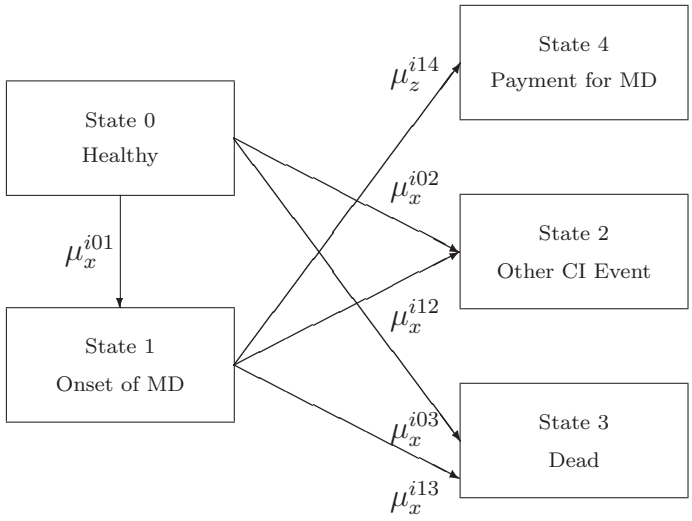

Figure 2.4: A semi-Markov model for myotonic dystrophy and critical illness insurance for insurance applicants with genotype g_i. The variable z is the duration since onset of myotonic dystrophy.

The intensities μ_x^{i01}, μ_x^{i02} and μ_x^{i03} in Figure 2.4 are the same as the corresponding intensities in Figure 2.3. See Section 2.3.1 for details. Since the intensity μ_z^{i14} is duration dependent, this model is semi-Markov. We assume in Figure 2.4 that $\mu_x^{i02} = \mu_x^{i12}$ and $\mu_x^{i03} = \mu_x^{i13}$.

We use the accelerated life model to represent the intensity μ_z^{i14} based on post-onset mortality $\mu_z^{MD,Mortality}$. Huntington's disease (HD) is fully studied in Gutiérrez & Macdonald (2004), in which the payment of a CI benefit for HD also occurs some time after onset. In this paper, the authors assume that the payment will be paid after 1/3 (Stage II) or 2/3 (Stage III) of the mean remaining lifetime after onset of HD (on average), which is governed by the post-onset mortality $\mu_z^{HD,Mortality}$. Stage II and Stage III represent two stages of the progression of HD, defined in Harper (1996). In this paper, we continue to use the terms Stage II and Stage III to represent the times at which an insurance company might pay a CI claim in respect of MD. However, unlike HD, these two stages do not carry any medical explanation, but simply their mathematical meaning. The accelerated lifetime model is used to deal with delayed payment. Please refer to Gutiérrez & Macdonald (2004) for more details about this model. We only give a simple description in Section 2.3.2.

The Accelerated Lifetime Model

The accelerated lifetime model is used in industry to test the lifetime of a component. The idea is that since we might not be able to wait for long enough to assess the lifetime of a component under normal circumstances, instead we assess this component's shortened lifetime when under pressure, stress, or heat. Then we use a parametric model to find the normal (i.e. usual) distribution of the lifetime of the component.

Given the distribution $F_X(x)$ of a random variable X representing a lifetime, we

multiply the timescale by a constant $\phi \geqslant 1$ to obtain a new random variable Y such that:

$$F_Y(x) = \mathrm{P}[X \leq \phi x] = F_X(\phi x). \tag{2.6}$$

Referring to Equation (2.1), the intensity $\mu_X(x)$ corresponding to the random variable X is defined as:

$$\mu_X(x) = \frac{F'_X(x)}{1 - F_X(x)}, \tag{2.7}$$

The relation between the intensities associated with X and Y is:

$$\mu_Y(x) = \phi \mu_X(\phi x), \tag{2.8}$$

the relation between the mean of the random variable Y and X is that:

$$\mathbf{E}[Y] = \frac{1}{\phi}\mathbf{E}[X] \tag{2.9}$$

An Application of the Accelerated Lifetime Model

In the case of HD, $\phi = 3$ and $\phi = 1.5$ were chosen to correspond to a claim being paid after 1/3 (Stage II) and 2/3 (Stage III) of the mean remaining lifetime after onset of HD, respectively (Gutiérrez & Macdonald, 2004). In the case of MD, we choose the same values of ϕ as suggested in Gutiérrez & Macdonald (2004). The intensity μ_z^{i14} is calculated based on post-onset mortality $\mu_z^{MD,Mortality}$ (Equation (2.5)). In the case that $\phi = 3$, we have:

$$\begin{aligned} \mu_z^{i14} &= 3 \times \frac{\exp\left(0.001903(3z)^3 - 0.06907(3z)^2 + 1.007(3z) - 6.082\right)}{1 + \exp\left(0.001903(3z)^3 - 0.06907(3z)^2 + 1.007(3z) - 6.082\right)} \\ &\quad \times (0.005709(3z)^2 - 0.1381(3z) + 1.007), \end{aligned} \tag{2.10}$$

and in the case that $\phi = 1.5$, we have:

$$\begin{aligned} \mu_z^{i14} &= 1.5 \times \frac{\exp\left(0.001903(1.5z)^3 - 0.06907(1.5z)^2 + 1.007(1.5z) - 6.082\right)}{1 + \exp\left(0.001903(1.5z)^3 - 0.06907(1.5z)^2 + 1.007(1.5z) - 6.082\right)} \\ &\quad \times (0.005709(1.5z)^2 - 0.1381(1.5z) + 1.007), \end{aligned} \tag{2.11}$$

The payment for MD should happen some time between the onset of MD and death. Two extreme scenarios are that:

(a) Payment for MD is made at onset, which implies that $\phi = +\infty$;

(b) Payment for MD is made at death, which implies that $\phi = 1$.

Our choices of $\phi = 3.0$ and $\phi = 1.5$ are two values between these two extreme values 1 and $+\infty$. Although, unlike in the case of HD, we can not locate any paper to substantiate these choices, 1/3 (Stage II) and 2/3 (Stage III) of the mean remaining lifetime after onset of MD are two appropriate points to choose to represent the time uncertainty of payment for MD. See Section 2.4.1 for a short discussion on the relation between ϕ and the resulting premium rates.

2.3.3 Numerical Methods

Once a Markov model has been fully calibrated, generally we have two questions which need to be answered, the first being how to find occupancy probabilities, and the second being how to find the reserves we need to set up at any time for a life present in any state before the policy expires. The first question is answered by solving the Kolmogorov forward equations and the second by solving Thiele's equations. These are both systems of linear ordinary differential equations (ODE).

Assuming we are looking at an N-state model, the states being labelled $0, 1, 2 \ldots (N-1)$. The transition rate from state i to j $(i \neq j)$ is defined as:

$$\mu_x^{ij} = \lim_{dt \to 0} \frac{{}_{dt}p_x^{ij}}{dt}, \tag{2.12}$$

where ${}_tp_x^{ij}$ is defined as:

$${}_tp_x^{ij} = \mathbf{P}\left[\text{in state } j \text{ at age } x+t \mid \text{ in state } i \text{ at age } x\right]. \tag{2.13}$$

Then Kolmogorov's Forward Equations are:

$$\frac{\partial}{\partial t}{}_tp_x^{ij} = \sum_{k\neq j} {}_tp_x^{ik}\mu_x^{kj} - \sum_{k\neq j} {}_tp_x^{ij}\mu_x^{jk}. \tag{2.14}$$

Our other major task is to calculate the prospective reserve set up by insurers in respect of an insured person at any time t. Two types of cash flows need to be defined.

(a) Define b_{x+t}^i as the continuous rate of payment from the insured person per annum while remaining in state i at age $x+t$.

(b) Define c_{x+t}^{ij} as the lump-sum payment to the insured person made when transition from state i to state j takes place at age $x+t$.

Further suppose there is a constant force of interest δ. Define ${}_tV_x^i$ to be the Expected Present Value (EPV) of the future loss conditional on being in state i at time t. Thiele's equations follow:

$$\frac{\partial}{\partial t}{}_tV_x^i = {}_tV_x^i\delta + b_{x+t}^i - \sum_{j\neq i} \mu_{x+t}^{ij}(c_{x+t}^{ij} + {}_tV_x^j - {}_tV_x^i) \tag{2.15}$$

Normally, we can only solve this system numerically. Fourth Order Runge-Kutta methods (Press *et al.*, 1988) are one of the most commonly used methods to solve systems of ODE. We used this method with step-size 0.0005 year. See Appendix D for technical details.

In Figure 2.4, the intensity μ_z^{i14} is duration dependent, which means this model is semi-Markov. Equations (2.14) and (2.15) are not valid. However, we can bring this semi-Markov model back to a Markov framework regime by assuming the insurer reinsures all risks once an insured person suffers onset of MD. Define ${}_{z,t}{}^iV_x^j$ as the prospective reserve for a life in state ij at age $x+t$ with duration z since entry into the state. So the insurers need only pay a single premium equal to the price of this reinsurance deal, which is, if onset occurs at time t:

$$ {}_{0,t}{}^{i}V_x^j = \int_0^{T-t} e^{-\delta s}\, {}_sp_{x+t}^{i11}(\mu_s^{i14} + \mu_{x+t+s}^{i12})ds, \tag{2.16} $$

where ${}_sp_{x+t}^{i11}$ is defined as:

$$ {}_sp_{x+t}^{i11} = \exp\left(-\int_0^s (\mu_h^{i14} + \mu_{x+t+h}^{i12} + \mu_{x+t+h}^{i13})dh\right), \tag{2.17} $$

and T is the original policy term.

2.4 Critical Illness Insurance Underwriting

2.4.1 Underwriting of an Applicant who is a Myotonic Dystrophy Mutation Carrier

Knowing a genetic test result, based on the model shown in Figure 2.4, we can calculate the level rate of premium for a unit sum assured for a healthy insurance applicant with different genotype, i.e. mutation CTG250+ and CTG250– carriers and non-mutation carriers. Tables 2.4 and 2.5 show the net level premium rates for different entry ages and policy terms, payable continuously, for a unit sum assured for both males and females, as a percentage of the premium for standard risk, assuming claims arise at Stage II or Stage III, respectively.

The 'standard risk' means healthy non-mutation carriers. Note that unhealthy non-mutation carriers might be either insured at an increased premium or become completely uninsurable. We observe some features in Tables 2.4 and 2.5.

(a) As suggested by Brackenridge & Elder (1998), applicants might be accepted at a rate of +200% to +300% for certain CI insurance products, i.e. 300% to 400% of the premiums for standard population, but most insurers would decline cases where the premium rating was over +200% to +250%, i.e. 300% to 350% of the premium

Table 2.4: Level net premium rates for carriers of MD mutations, as a percentage of the premium for standard risks. Claims arising at Stage II ($\phi = 3.0$). Critical illness insurance contracts with unit sum assured.

Male				Female			
Age (Years)	Term (Years)	CTG250– %	CTG250+ %	Age (Years)	Term (Years)	CTG250– %	CTG250+ %
20	10	1,512	7,378	20	10	921	4,333
	20	1,517	6,891		20	1,079	4,791
	30	1,037	3,785		30	895	3,224
	40	689	2,203		40	691	2,198
30	10	939	4,547	30	10	735	3,467
	20	821	3,075		20	757	2,810
	30	582	1,863		30	612	1,962
40	10	474	1,517	40	10	477	1,528
	20	431	1,162		20	480	1,317
50	10	260	618	50	10	301	752

for standard population. Therefore, almost all the insurance applicants with mutations, either CTG250+ or CTG250– become uninsurable, with few exceptions. The premium charged for healthy mutation carriers could be as high as 74 times standard premium (male mutation CTG250+ carrier, entry age 20, policy term 10, claim arising at Stage II).

(b) In Table 2.4, the exceptions are male and female mutation CTG250– carriers at age 50 purchasing 10-year term insurance. They might be accepted at a higher premium rate (i.e. above +200%), although in practice they might be quite likely

Table 2.5: Level net premium rates for carriers of MD mutations, as a percentage of the premium for standard risks. Claims arising at Stage III ($\phi = 1.5$). Critical illness insurance contracts with unit sum assured.

Male				Female			
Age (Years)	Term (Years)	CTG250– %	CTG250+ %	Age (Years)	Term (Years)	CTG250– %	CTG250+ %
20	10	551	2,464	20	10	376	1,475
	20	805	3,819		20	644	2,670
	30	610	2,485		30	620	2,124
	40	420	1,486		40	517	1,488
30	10	350	1,521	30	10	314	1,176
	20	432	1,782		20	477	1,634
	30	329	1,170		30	441	1,236
40	10	197	550	40	10	227	554
	20	230	690		20	322	780
50	10	130	264	50	10	167	307

to be declined. In Table 2.5, both male and female mutation CTG250– carriers at ages 40 and 50 purchasing 10-year term insurance are likely to be accepted at an extra premium rates.

(c) At lower ages, females are charged a higher level of standard premium, while at higher ages, males are charged a higher level of standard premium. Since the onset rate of MD and the post-onset mortality rate are not gender-differentiated, these differences come from intensities μ_x^{i02} and μ_x^{i03}, which represent the intensity of contracting other CI and adjusted intensity of death respectively (See Appendix A for details).

(d) Mutation CTG250+ carriers have a higher onset rate than mutation CTG250– carriers. Therefore the corresponding premium rates charged for the former are higher.

(e) The extra premium rates are very sensitive to the entry age. The relative level of extra premium rate decreases with increased policy term and increased entry age.

(f) We can see in Table 2.4 and 2.5 that the premium rates for mutation carriers are very sensitive to the choice of ϕ. In Section 2.3.2, we discussed that the value of ϕ ranges between 1 and $+\infty$: the former represents the scenario that the payment for MD is made at death and the latter represents the scenario that the payment for MD is made at onset. In order to see the relation between ϕ and the resulting premiums, we choose ϕ to be $1, 1.2, 1.5, 2, 3$ or 6 (corresponding to a claim being paid after $1, 5/6, 2/3, 1/2, 1/3$ or $1/6$ of the mean remaining lifetime after onset of MD. For simplicity, we calculate the net level premium rates as a percentage of the premium for standard risks for both males and females aged 30 seeking 30-year policies. These percentages are plotted against ϕ in Figure 2.4.1.

We can see some features in Figure 2.4.1:

(a) The premium rates increase dramatically with ϕ, especially for mutation CTG250+ carriers.

(b) Even we choose $\phi = 1$, which implies that the payment for MD is made at death and gives the lowest premium rates, mutation carriers are still very likely to be rejected by the insurers for both males and females.

2.4.2 Underwriting Based on Family History Only

Based on the Mendelian rule, a new-born baby with a family history, meaning that one of the baby's parents or siblings is a mutation carrier, has a 50% probability of being a

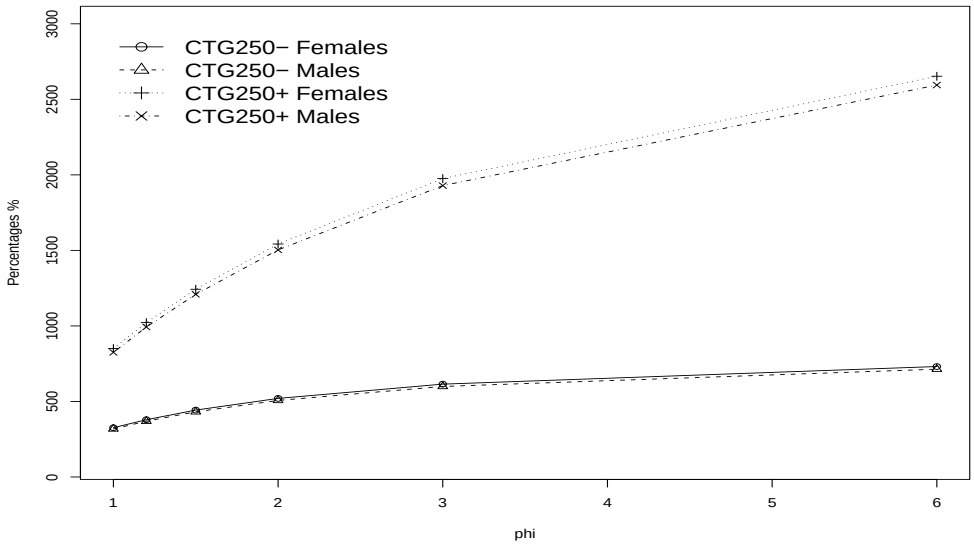

Figure 2.5: The net level premium rates for both males and females aged 30 seeking 30 years CI policies, calculated with various choices of ϕ, presented as a percentage of the premium for standard risks.

mutation carrier, assuming there is only a very small chance that both parents are mutation carriers. Recall that in Section 2.2.4 we estimated that mutation carriers have equal chances to carry mutation CTG250+ and mutation CTG250– at birth. We also assume that we cannot infer from the parent's age at onset whether a CTG250– or CTG250+ mutation is implicated. Therefore at birth, a new born baby has a 25% chance to be a mutation CTG250+ carrier, a 25% chance to be a mutation CTG250– carrier, and a 50% chance to be a non-mutation carrier (NC). We use p_i to represent these probabilities for genotype g_i. This is the start of our calculation. Let $\mathbf{OP}_x^{i0}$ be the probability of remaining in State 0 by age x for genotype g_i, according to the model in Figure 2.3. Please see Table 2.6 for numerical results.

Knowing the family history when underwriting, insurers are interested in the proba-

Table 2.6: Numerical results of occupancy probability $\mathbf{OP}_x^{i0}$ for both males and females with genotype g_i at different ages.

	Females				Males		
Age	NC	CTG250–	CTG250+	Age	NC	CTG250–	CTG250+
20	0.497750	0.233880	0.217254	20	0.496331	0.233213	0.216634
30	0.496036	0.207336	0.104627	30	0.492044	0.205667	0.103784
40	0.492684	0.162441	0.026150	40	0.486267	0.160325	0.025809
50	0.484113	0.111015	0.005400	50	0.473280	0.108530	0.005279

bilities:

$$\mathbf{P}\left[\text{genotype } g_i \mid \text{healthy at age } x \text{ and with family history}\right],$$

denoted $\mathbf{P}_i$, which can be calculated as:

$$\mathbf{P}_i = \frac{\mathbf{OP}_x^{i0} p_i}{\sum_{j=1}^{3} \mathbf{OP}_x^{j0} p_j}. \tag{2.18}$$

Please see Table 2.7 for the numerical results of probability $\mathbf{P}_i$.

Table 2.7: Numerical results of probabilities P_i for males and females with genotype g_i at different ages.

	Females				Males		
Age	NC	CTG250–	CTG250+	Age	NC	CTG250–	CTG250+
20	0.524564	0.246479	0.228957	20	0.524564	0.246479	0.228957
30	0.613907	0.256604	0.129489	30	0.613907	0.256604	0.129489
40	0.723179	0.238437	0.038384	40	0.72318	0.238437	0.038383
50	0.806145	0.184862	0.008992	50	0.806146	0.184862	0.008992

In this case, the net level premium will be the ratio of the EPV of 1 unit of benefit to the EPV of 1 unit of premium each weighted by the $\mathbf{P}_i$. Table 2.8 shows the net level premiums as a percentage of the standard rate, payable continuously, for males and females with a family history, claims arising at Stage II and III for different entry age and policy term. In Table 2.8, we can see some features as follows:

(a) In most cases, insurance applicants with a family history are uninsurable. The level of extra premiums charged can be as high as 20 times the standard premium (male, entry age 20, policy term 10 years, claims arising at stage II). However, more people at higher ages can be accepted at an increased premium rates (male and female, entry age 50, claims arising at either stage II or III).

(b) Compared with the results from Tables 2.4 and 2.5, the level of extra premiums decreases. This is the effect of averaging out between 3 sub-populations, and especially because non-mutation carriers carry higher weights than the mutation carriers. In particular, if an applicant at risk of MD remains healthy at a high age, they become more unlikely to be a mutation carrier, and therefore their risks decrease to an acceptable level.

(c) As in Tables 2.4 and 2.5, the extra premiums are more sensitive to age than to policy term.

(d) We compare the results in Tables 2.4, 2.5 and 2.8. In all cases, the family history underwriting makes the premiums for mutation CTG250+ carriers significantly lower than the corresponding cases under the genetic test results underwriting. For mutation CTG250– carriers, this is only true at late ages, but not true at early ages. This is generally because the weights assigned to each genotypes, i.e. probabilities P_i, play a strong role in family history underwriting. In Table 2.7, the possibilities

of being mutation carriers are still high at early ages, but decrease dramatically at late ages, especially for mutation CTG250+ carriers.

Table 2.8: Level net premium rates for critical illness insurance applicant with family history of MD, as a percentage of the premium for standard risks. Critical illness insurance contracts with unit sum assured.

Male				Female			
Age	Term	Claim Arising At Stage II(%)	Claim Arising At Stage III(%)	Age	Term	Claim Arising At Stage II(%)	Claim Arising At Stage III(%)
20	10	2,046	743	20	10	1,231	478
	20	1,746	1,052		20	1,237	776
	30	954	680		30	823	616
	40	575	427		40	574	451
30	10	827	342	30	10	651	289
	20	551	371		20	511	364
	30	356	265		30	370	294
40	10	234	141	40	10	235	147
	20	199	153		20	213	172
50	10	132	108	50	10	141	114

2.5 Critical Illness Insurance and Adverse Selection

2.5.1 Adverse Selection and Moratoria on Genetic Information in Underwriting

Adverse selection may arise in the context of genetics and insurance if persons who know they are at increased risk because they have some genetic information (see Chapter 1 for details) are more likely to buy insurance and need not share that information with the insurers. Governments may impose moratoria on insurers regarding the use of all types of genetic information. In the UK, there is currently an agreed moratorium on the use of DNA-based test results (a narrow definition of genetic information), while in Sweden there is a moratorium covering both DNA-based test results and family history (a broad definition of genetic information).

In the underwriting process, insurers tend to pool applicants with roughly equal levels of risk into underwriting classes and charge the same premium within each underwriting class. However, when a moratorium is imposed, applicants at different level of risks might be forced into the same underwriting class. In this case, applicants knowing their increased level of risk would buy more insurance, and a premium set according to the average risk in the underwriting class will not be sufficient to cover the actual claims. Insurers would have to increase the premium rate in order to recover the loss caused by the behaviour of 'adverse selectors'. The increase in the premium rate is the cost of adverse selection, which will depend on the type of insurance, the size of the market for that insurance and the form of moratorium imposed, as well as on the epidemiology of each genetic disorder of interest.

We use the CI insurance market model shown in Figure 2.6 to model the life history of a person. We are interested in the relevant events in a CI insurance market, e.g. purchase of the insurance product, taking a genetic test, etc. The CI insurance market model is

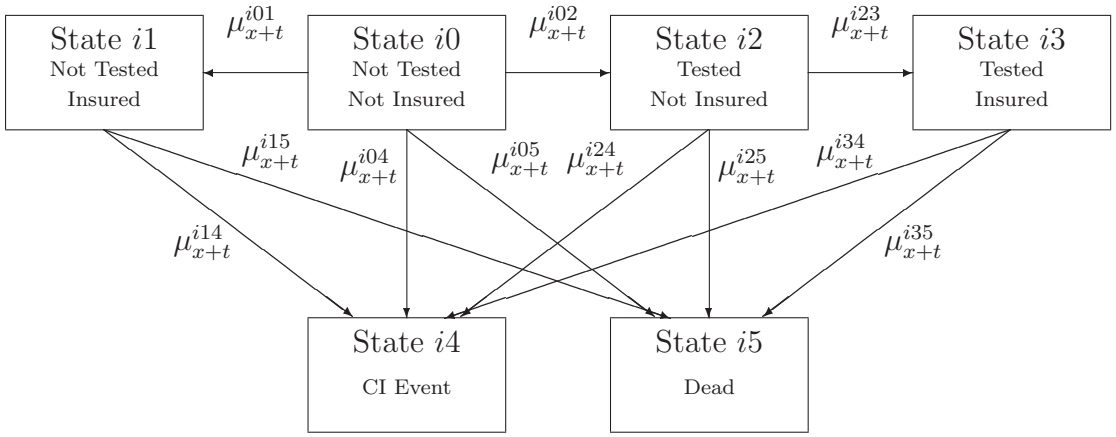

Figure 2.6: A Markov model of insurance purchase and CI insurance events for a person in sub-population i.

adopted from Gutiérrez & Macdonald (2004). This is a continuous-time discrete-state Markov model.

This model is deceptively simple in the case of MD, but is good for purposes of demonstration. A more detailed expanded model for computation is displayed in Appendix B for comparison. Label i indicates the sub-populations. Here we have 5 sub-populations as follows:

(a) sub-population 1: Non-mutation carriers, with no family history.

(b) sub-population 2: Carrying mutation CTG250+, with family history.

(c) sub-population 3: Not carrying mutation CTG250+, with family history.

(d) sub-population 4: Carrying mutation CTG250–, with family history.

(e) sub-population 5: Not carrying mutation CTG250–, with family history.

The proportion of the population in sub-population i, at birth, is denoted p_i. Given our assumed prevalence rate of MD in the UK of 7.1 per 100,000, we take proportions in each sub-population to be as follows: $p_1 = 0.999858$, $p_2 = p_3 = p_4 = p_5 = 3.55 \times 10^{-5}$.

The details of this model have been extensively discussed in Gutiérrez & Macdonald (2003). Therefore we only introduce the main features of this model briefly here:

(a) We assume that the CI insurance market operates between ages 20 and 60. Individuals could buy insurance before or after taking a genetic test. They could also become uninsurable because of onset of MD, another critical illness or death.

(b) The insurer classifies the whole population into underwriting classes according to the risk factors used. Everyone in an underwriting class is charged the same rate of premium, which we assume satisfies the equivalence principle within that class. If there is no moratorium, insurers will categorize the population according to the heterogeneity of risks. However, when a moratorium is imposed, very different levels of risk will be forced into the same underwriting class. In the UK, family history may be used in the underwriting process, while genetic test results usually may not. So individuals could have better information about their risk status than the insurance company if they have had a genetic test. When people are charged lower premiums than they would have been had they disclosed their true risk, adverse selection arises. In Sweden, even family history may not be used in underwriting. In this case, the level of adverse selection should in theory be higher.

(c) The rate of premium within each underwriting class is calculated according to the equivalence principle. However, we do not use level premiums, because a level premium depends on the age x at which insurance was purchased and policy term. This is not compatible with Thiele's equations. Therefore, we need to find an age-dependent premium rate. This rate is the weighted average intensity from the

insured states in a given underwriting class to the CI claim states. If we use $\mathcal{C}$ to denote an underwriting class, the rate of premium in that class, denoted as $\rho^{\mathcal{C}}_{x+t}$, is calculated as follows:

$$\rho^{\mathcal{C}}_{x+t} = \frac{\sum_{ij \in \mathcal{C}} p_{i\ t} p^{i0j}_{x} \mu^{ij4}_{x+t}}{\sum_{ij \in \mathcal{C}} p_{i\ t} p^{i0j}_{x}}, \tag{2.19}$$

where label ij stands for the state j shown in Figure 2.6 for sub-population i. The delayed payment also need to be allowed for by using Equation (2.16). Once everything is brought back to a Markov framework regime, we can use Thiele's equations to calculate the expected insurance loss incurred conditional on occupying any state in each sub-population. The insurance loss is defined as:

$$\text{Insurance Loss} = \text{Present Value of Outgo} - \text{Present Value of Income} \tag{2.20}$$

We ignore the expenses in this thesis. So 'outgo' stands for insurance payment and 'income' stands for premium from an insurer's perspective.

(d) The sum of these insurance losses conditional on being in state $i0$, weighted by the prevalence rate of each sub-population, is the insurance loss in the whole insurance market.

(e) If there is no adverse selection, the expected insurance loss in the whole insurance market is zero, i.e. EPV of loss without adverse selection = 0, because the equivalence principle is correctly applied. However, if there is adverse selection, the insurance loss is non-zero. Insurers will have to increase all premiums by the proportion:

$$\frac{\text{EPV of loss with adverse selection} - \text{EPV of loss without adverse selection}}{\text{EPV of premium payable with adverse selection}}, \tag{2.21}$$

to absorb the loss. This quantity is our measure of the level of adverse selection.

(f) The risk measure of the cost of adverse selection varies when different models are chosen. For example, in de Jong & Ferris (2006), the authors used correlation between the amount of insurance required and the risk premium rate for an individual within a particular risk group to model the cost of adverse selection in the context of insurance. Adverse selection has also been studied from an economist's perspective, e.g. in Hoy & Polborn (2000), the authors constructed an economic model and used the expected utility to evaluate the social value of genetic information. Our choice of risk measure is suitable in this context when we use Markov model to model the life histories and policyholder behaviours in the insurance market.

2.5.2 Calibration of the CI Market Model

In the CI insurance market model in Figure 2.6, several intensities have been fully defined, including those into CI claim states and into the dead state (See Appendix A for details). Several other intensities need careful explanations, e.g. insurance purchase rates and genetic testing rates.

(a) Insurance purchase rates μ^{i01}_{x+t} and μ^{i23}_{x+t}. The function of the purchase rate is two-fold. First, we use the purchase rate in sub-population 1 (we call it normal rate) to represent the size of insurance market. If the purchase rate is equal to 0.05 per annum, we call this a large market, while if the purchase rate is equal to 0.01 per annum, we call this a small market. These assumptions implies that about 85% and 30% of persons, respectively, will buy insurance at some time between

ages 20 and 60. Second, in a large market, sub-populations 2 – 5 might choose to purchase CI contracts at normal rate, or half or normal rate, or not at all, while in a small market, sub-populations 2 – 5 do not purchase any CI contract. The nil purchase rate assumption is probably more realistic because the applicants with family history of MD are normally rejected by the insurers or accepted at an extremely high premium rate. (See Table 2.8 for details.)

(b) Genetic testing rate μ^{i02}_{x+t}. The baseline test rate is 0.014 per annum for ages 20 – 40, which implies that about 10% would be tested after 8 years. The other two scenarios are 0.014 per annum for ages 20 – 60 (testing longer period of genetic test), and 0.035 per annum for ages 20 – 40 (implying about 24% would be tested in 8 years).

The form of adverse selection includes:

(a) Increased purchase rate. In the case of moderate adverse selection, we assume the purchase rate is double that of the normal rate, i.e. 0.1 per annum in large market and 0.02 per annum in small market. These assumption imply that 63% and 18% of population would buy insurance in 10 years in large and small market, respectively. In the case of severe adverse selection, we assume the purchase rate is 0.25 per annum. This assumption is deliberately high, which implies that 91% of population would buy insurance in 10 years in both large and small market.

(b) Increased sum assured. In (a) we assume that 'adverse selectors' buy the same amount of insurance as normal. This possibility is the second source of adverse selection, e.g. increased sum assured. Gutiérrez & Macdonald (2004) suggest that in this case the general premium increase, e.g. the cost of adverse selection, is very nearly proportionate to the multiple of the average sum assured taken out by 'adverse selectors', therefore we omit the tables.

2.5.3 Moratoria on Using Genetic Test Results

When moratoria on using genetic test results are imposed, either on all test results or on negative test results, there are two underwriting classes (UC), that we call UC-I and UC-II. In UC-I, persons will be charged standard premiums and in UC-II, persons will be charged extra premiums because they have family histories. Tables 2.9 and 2.10 display the cost of moderate adverse selection following a moratorium on the use of genetic test results, with family history underwriting allowed, in a CI insurance market operating between ages 20 and 60, assuming the claims for MD will be paid at Stage II or Stage III, respectively. Tables 2.11 and 2.12 demonstrate the results of severe adverse selection. Generally, the increases are all very small, even with severe adverse selection, but we can observe the following features:

(a) The relative costs are more substantial in the smaller market.

(b) Premium increases are lower if the moratorium applies only to adverse test results. This is because tested persons who are not mutation carriers will be removed from the underwriting class rated for family history and charged standard premiums. The latter then contains a higher proportion of mutation carriers, so the premium charged in respect of that class is higher.

(c) A longer period of genetic testing has little effect on the cost of adverse selection. This is because of high penetrance of MD mutations. Only very few applicants carrying a MD mutation will remain healthy after age 40.

(d) The cost of adverse selection is higher in the case that claims arise at Stage II than when claims arise at Stage III.

Table 2.9: Percentage increases in CI insurance premium rates arising from moderate adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. CI Market operating between ages 20 and 60. Claims arising at Stage II of MD.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Males	Females	Males	Females
				%	%	%	%
Large	Normal	0.014	20–40	0.00090	0.00096	0.00086	0.00092
	Half	0.014	20–40	0.00184	0.00195	0.00175	0.00185
	Nil	0.014	20–40	0.00561	0.00594	0.00558	0.00591
Small	Nil	0.014	20–40	0.00529	0.00571	0.00526	0.00568
Large	Normal	0.014	20–60	0.00092	0.00097	0.00088	0.00093
	Half	0.014	20–60	0.00188	0.00199	0.00179	0.00189
	Nil	0.014	20–60	0.00590	0.00623	0.00587	0.00620
Small	Nil	0.014	20–60	0.00552	0.00595	0.00549	0.00592
Large	Normal	0.035	20–40	0.00204	0.00216	0.00185	0.00196
	Half	0.035	20–40	0.00412	0.00436	0.00367	0.00389
	Nil	0.035	20–40	0.01236	0.01308	0.01229	0.01301
Small	Nil	0.035	20–40	0.01169	0.01262	0.01163	0.01256

Table 2.10: Percentage increases in CI insurance premium rates arising from moderate adverse selection. Moratoria on the use of all genetic test results, family history underwriting still allowed. CI Market operating between ages 20 and 60. Claims arising at Stage III of MD.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Males %	Females %	Males %	Females %
Large	Normal	0.014	20–40	0.00067	0.00071	0.00064	0.00068
	Half	0.014	20–40	0.00133	0.00141	0.00125	0.00133
	Nil	0.014	20–40	0.00423	0.00451	0.00426	0.00450
Small	Nil	0.014	20–40	0.00397	0.00427	0.00396	0.00426
Large	Normal	0.014	20–60	0.00067	0.00071	0.00064	0.00068
	Half	0.014	20–60	0.00133	0.00141	0.00125	0.00133
	Nil	0.014	20–60	0.00445	0.00469	0.00444	0.00468
Small	Nil	0.014	20–60	0.00412	0.00442	0.00411	0.00441
Large	Normal	0.035	20–40	0.00153	0.00163	0.00137	0.00145
	Half	0.035	20–40	0.00301	0.00320	0.00264	0.00281
	Nil	0.035	20–40	0.00942	0.00996	0.00940	0.00994
Small	Nil	0.035	20–40	0.00879	0.00947	0.00877	0.00945

Table 2.11: Percentage increases in CI insurance premium rates arising from severe adverse selection. Moratoria on the use of all genetic test results, family history underwriting still allowed. CI Market operating between ages 20 and 60. Claims arising at Stage II of MD.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Males %	Females %	Males %	Females %
Large	Normal	0.014	20–40	0.00214	0.00227	0.00206	0.00218
	Half	0.014	20–40	0.00329	0.00348	0.00314	0.00333
	Nil	0.014	20–40	0.00817	0.00866	0.00813	0.00861
Small	Nil	0.014	20–40	0.02462	0.02670	0.02450	0.02656
Large	Normal	0.014	20–60	0.00219	0.00232	0.00210	0.00222
	Half	0.014	20–60	0.00338	0.00358	0.00323	0.00342
	Nil	0.014	20–60	0.00866	0.00915	0.00861	0.00911
Small	Nil	0.014	20–60	0.02609	0.02823	0.02594	0.02808
Large	Normal	0.035	20–40	0.00485	0.00513	0.00444	0.00470
	Half	0.035	20–40	0.00735	0.00778	0.00665	0.00704
	Nil	0.035	20–40	0.01794	0.01902	0.01785	0.01892
Small	Nil	0.035	20–40	0.05407	0.05865	0.05379	0.05833

Table 2.12: Percentage increases in CI insurance premium rates arising from severe adverse selection. Moratoria on the use of all genetic test results, family history underwriting still allowed. CI Market operating between ages 20 and 60. Claims arising at Stage III of MD.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Males %	Females %	Males %	Females %
Large	Normal	0.014	20–40	0.00168	0.00178	0.00162	0.00172
	Half	0.014	20–40	0.00250	0.00266	0.00239	0.00254
	Nil	0.014	20–40	0.00627	0.00665	0.00626	0.00664
Small	Nil	0.014	20–40	0.01890	0.02052	0.01887	0.02047
Large	Normal	0.014	20–60	0.00168	0.00176	0.00161	0.00171
	Half	0.014	20–60	0.00251	0.00267	0.00239	0.00254
	Nil	0.014	20–60	0.00659	0.00696	0.00657	0.00695
Small	Nil	0.014	20–60	0.01985	0.02147	0.01981	0.02142
Large	Normal	0.035	20–40	0.00383	0.00407	0.00351	0.00373
	Half	0.035	20–40	0.00566	0.00601	0.00510	0.00541
	Nil	0.035	20–40	0.01380	0.01465	0.01377	0.01461
Small	Nil	0.035	20–40	0.04159	0.04517	0.04151	0.04507

2.5.4 Moratoria on Using Genetic Test Results and Family History

A moratorium on using genetic test results and family history causes premiums to increase in two different ways.

(a) Persons in the higher-risk sub-population now could purchase a normal amount of insurance cover at the ordinary rate. A new underwriting class forms, comprising persons in all sub-populations. This will increase the premium rate. But this is arguably not adverse selection if the behaviour of persons in the high-risk sub-populations is just the same as that of persons in the low-risk sub-populations.

(b) Further, insurance buyers may increase their purchase rate and/or amount of cover in reaction to the information they have and the lower premiums they have been charged. In this way, adverse selection arises.

Tables 2.13 and 2.14 show the cost of adverse selection when a moratorium on all genetic information is in place, assuming the claim for MD arises at Stage II or Stage III, respectively. Both moderate and severe levels of adverse selection are shown. The rate of genetic testing is 0.014 per annum with moderate adverse selection, and 0.035 per annum betweeen ages 20 and 40 with severe adverse selection. We can observe that in absolute terms these increases are negligible.

Table 2.13: Percentage increases in standard premium rates for CI insurance arising from new underwriting classes, and in all premiums arising from moderate or severe adverse selection, following a moratorium on the use of all genetic test results and family history. CI market operating between ages 20 and 60. Claims arising at Stage II of MD.

	OR Premium Increase Arising From New		Premium Increase Arising From		Premium Increase Arising From	
Size of	Underwriting Classes		Moderate Adverse Selection		Severe Adverse Selection	
Market	Males	Females	Males	Females	Males	Females
	%	%	%	%	%	%
Large	0.03173	0.03357	0.01635	0.01726	0.03602	0.03800
Small	0.02520	0.02731	0.02160	0.02339	0.17898	0.19337

Table 2.14: Percentage increases in standard premium rates for CI insurance arising from new underwriting classes, and in all premiums arising from moderate or severe adverse selection, following a moratorium on the use of all genetic test results and family history. CI market operating between ages 20 and 60. Claims arising at Stage III of MD.

	OR Premium Increase Arising From New		Premium Increase Arising From		Premium Increase Arising From	
Size of	Underwriting Classes		Moderate Adverse Selection		Severe Adverse Selection	
Market	Males	Females	Males	Females	Males	Females
	%	%	%	%	%	%
Large	0.02393	0.02536	0.01271	0.01344	0.02828	0.02988
Small	0.01873	0.02034	0.01621	0.01758	0.13862	0.15003

2.6 A Life Insurance Model

In Section 2.2.3, we found that the mortality after onset of MD depends on duration since onset as well as age. Therefore, for a life insurance policy, we use a semi-Markov model adopted from Gutiérrez & Macdonald (2004) shown in Figure 2.7, to calculate net level life insurance premium rates.

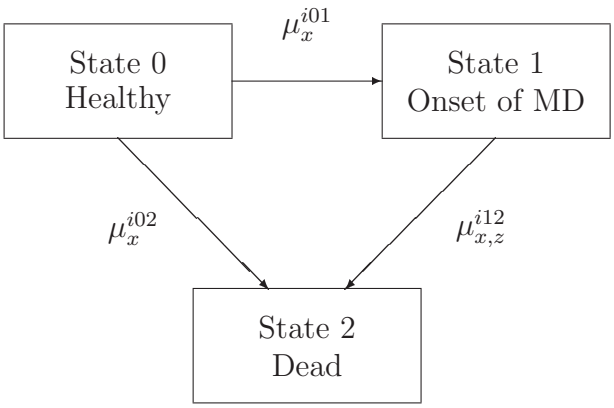

Figure 2.7: A multiple-state semi-Markov model for myotonic dystrophy and life insurance for insurance applicants with genotype g_i. Mortality after onset depends on age x and duration since onset z.

Intensity μ_x^{i01} follows the definition of the onset rate of MD, which is specified in Section 2.2.2. For intensity μ_x^{i02}, because the impact of MD on mortality is negligible, so we use the life table ELT15 without any adjustment. Post-onset mortality $\mu_{x,z}^{i12}$ is defined in Section 2.2.3, i.e.

$$\mu_{x,z}^{i12} = \mu_{x,z}^{MD,Dead}. \tag{2.22}$$

In this model, insured applicants pay premiums continuously while remaining in states

0 and 1, and the claim is paid when they enter state 2 either directly from state 0, or through state 1. We calculate the net level premium for a unit sum assured. We assume a constant force of interest $\delta = 0.05$.

We can easily write out expressions for the EPVs of a unit of premium and a unit of benefit as follows, assuming the entry age is x and the policy term is n:

$$\text{EPV [Premium]} = \int_x^{x+n} e^{-\delta t}{}_t p_x^{i00} dt + \int_x^{x+n} e^{-\delta t}{}_t p_x^{i00} \mu_{x+t}^{i01} \int_0^{n-t} e^{-\delta s}{}_s p_{x+t,0}^{i11} ds dt, \quad (2.23)$$

$$\text{EPV [Benefit]} = \int_x^{x+n} e^{-\delta t}{}_t p_x^{i00} \mu_{x+t}^{i02} dt + \int_x^{x+n} e^{-\delta t}{}_t p_x^{i00} \mu_{x+t}^{i01} \int_0^{n-t} e^{-\delta s}{}_s p_{x+t,0}^{i11} \mu_{x+t+s,s}^{i12} ds dt. \quad (2.24)$$

Then the net level premium for a unit of benefit payable continuously should be EPV [Benefit] /EPV [Premium].

2.7 Life Insurance Underwriting

2.7.1 Underwriting Based on Myotonic Dystrophy Mutation Status

Table 2.15 shows the net level life insurance premium rates for mutation CTG250+ and CTG250– carriers as a percentage of those for non-mutation carriers, as the 'standard' premium rates.

(a) Generally, females' mortality is lower than that of males. Therefore premium increases are higher for females, especially for cover expiring at later ages.

(b) The premium increases are lower than those for CI insurance.

(c) As suggested by Brackenridge & Elder (1998), applicants might be accepted at a rate of +200% to +300% for certain life insurance products, but most insurers would decline cases where the premium rating was over +200% to +%250. We may see that male mutation CTG250– carriers at high ages are more likely to be insured at an increased rate than are females, while mutation CTG250+ carriers are generally uninsurable, with a few exceptions at higher ages.

Table 2.15: Level net premium rates for life insurance, known MD mutation carriers, as a percentage of the premium for standard risks.

Male				Female			
Age (Years)	Term (Years)	CTG250– %	CTG250+ %	Age (Years)	Term (Years)	CTG250– %	CTG250+ %
20	10	182	506	20	10	317	1,171
	20	365	1,311		20	678	2,739
	30	491	1,659		30	844	3,058
	40	445	1,264		40	730	2,203
30	10	234	781	30	10	345	1,346
	20	387	1,254		20	582	2,025
	30	394	1,102		30	600	1,782
40	10	189	426	40	10	246	629
	20	244	544		20	348	854
50	10	127	186	50	10	154	271

2.7.2 Underwriting Based on Family History Only

Table 2.16 shows the net level premium rates for insurance applicants with a family history, as a percentage of those for non-mutation carriers, as the 'standard' premium rates. These premium rates are the weighted average premium rates conditional on insurance applicants aged x being non-mutation carriers, mutation CTG250+ carriers and CTG250– carriers. The weights are the probabilities defined in equation (2.18).

(a) Females are still charged higher relative premium rates than males.

(b) Males are more likely to be insured at an increased premium rate, except cases in which applicants aged 20 seek a policy with term longer than 10 years. Females are only likely to be insured, at an increased premium rate, at high ages.

Table 2.16: Level net premium rates for a healthy applicant with a family history of MD, expressed as a percentage of the premium for standard risks.

Age (Years)	Term (Years)	Females %	Males %
20	10	397	213
	20	823	432
	30	894	518
	40	663	412
30	10	328	221
	20	453	311
	30	408	283
40	10	155	133
	20	184	149
50	10	111	106

2.8 Life Insurance and Adverse Selection

The cause of adverse selection and measurement of the cost of adverse selection was discussed in Section 2.5. We use a semi-Markov model shown in Figure 2.8 to model the life insurance market, in order to calculate the cost of adverse selection. This model is adopted from Gutiérrez & Macdonald (2004). Sub-populations and their prevalence rates are defined in Section 2.5.1. State $i6$ is labelled "Dead, or MD and Not Insured". If someone develops MD or dies before they buy life insurance, they will become uninsurable. The reasons for having two states labelled MD (after purchase of insurance) are as follows:

(a) These two states might sometimes be in two different underwriting classes.

(b) We wish to consider the cost of adverse selection, depending on any moratorium in force. The losses conditional being in state $i3$ and state $i5$ could be different, because adverse selection could include the purchase of larger amounts of insurance.

2.8.1 Calibration of the Life Insurance Market Model and Methodology of Calculating the Cost of Adverse Selection

Onset rates, mortality rates and post-onset mortality rates have been discussed in Section 2.2. The functions of the purchase rate and test rate are fully introduced in Section 2.5.2. The mechanism of calculating the cost of adverse selection in the CI insurance market has been detailed in Section 2.5.1. To calculate the cost of adverse selection in the life insurance market, based on the model shown in Figure 2.8, we will go through the following steps.

(a) As explained in Section 2.5.1, instead of using level net premiums, we calculate a premium rate which only depends on age. In a certain underwriting class $\mathcal{C}$, this

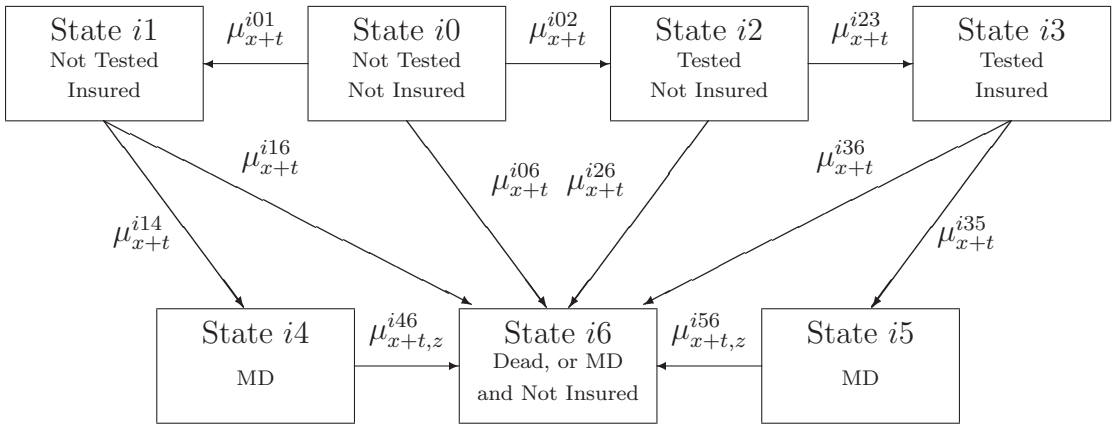

Figure 2.8: A semi-Markov model of insurance purchase and life insurance events for a person in sub-population i.

premium rate is equal to the weighted average intensity from the insured states into the claim states as follows:

$$\rho^{\mathcal{C}}_{x+t} = \frac{\sum_{ij\in\mathcal{C}} p_i \left({}_tp^{i0j}_x \mu^{ij6}_{x+t} + \int_0^t {}_{t,z}p^{i0j}_x \mu^{ij6}_{x+t,z} dz \right)}{\sum_{ij\in\mathcal{C}} p_i \left({}_tp^{i0j}_x + \int_0^t {}_{t,z}p^{i0j}_x dz \right)}, \tag{2.25}$$

where label ij stands for an insured state j shown in Figure 2.8 for sub-population i. The intensity $\mu^{ij6}_{x+t,z}$ in Equation (2.25) is duration dependent, so the probability ${}_{z,t}p^{i0j}_x$, where $j = 4, 5$, should indicate the duration this person remains in State 4 or 5. The probability ${}_{z,t}p^{i04}_x$ is defined as follows:

$$\begin{aligned} {}_{z,t}p^{i04}_x &= {}_{t-z}p^{i01}_x \times \mu^{i14}_{x+t-z} \times {}_zp^{i44}_{x+t-z} \\ &= {}_{t-z}p^{i01}_x \times \mu^{i14}_{x+t-z} \times \exp\left(- \int_0^z (\mu^{i46}_{x+t-z+y,y}) dy \right), \end{aligned} \tag{2.26}$$

which implies that this person enters into State 4 at age $x+t-z$ and remains in State 4 for duration z until (s)he reaches the age $x+t$.

The calculation of ${}_{z,t}p_x^{i05}$ is similar.

(b) We use the trick we introduced in Section 2.3.3 to bring the calculation back to a Markov model regime, by 'paying' a 'sum assured' equal to the policy value on entering an MD state from an insured state. The policy value ${}_{t,0}V_x^{i4}$ for a policy with term n years and sum assured 1 is:

$$ {}_{t,0}V_x^{i4} = \int_0^{n-t} e^{-\delta s} {}_s p_{x+t,0}^{i44} (\mu_{x+t+s,s}^{i46} - \rho_{x+t+s}^{\mathcal{C}}) ds, \tag{2.27} $$

and ${}_{t,0}V_x^{i5}$ is similar. Then we can use Thiele's equations to calculate the expected insurance loss conditional on being in any of the states.

(c) The expected insurance loss within the life insurance market is calculated as the weighted sum of insurance losses in each sub-population conditional on the probabilities of being in the starting states $i0$. Then the cost of adverse selection in the life insurance market is measured by the ratio:

$$ \frac{\text{EPV of loss with adverse selection} - \text{EPV of loss without adverse selection}}{\text{EPV of premiums payable with adverse selection}}. \tag{2.28} $$

2.8.2 Moratoria on Using Genetic Test Results

Table 2.17 shows results of moderate adverse selection in the life insurance market for both males and females, following a life insurance market operating between ages 20 and 60. As in CI market, we set the genetic testing rate at 0.014 per annum between ages 20 and 40 as our baseline, and compare it with another two scenarios, 0.014 per annum

between ages 20 and 60, and 0.035 per annum between ages 20 and 40. Table 2.18 shows the results of severe adverse selection in life insurance market. Here, we only consider a test rate of 0.035 per annum. All results, even in the case of severe adverse selection, are less than 0.1%, and so the cost of adverse selection may be regarded as negligible with respect to MD alone.

Table 2.17: Percentage increases in life insurance premium rates arising from moderate adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. Life insurance market operating between ages 20 and 60.

	Insurance Purchase		Age	Moratorium on Using			
Size of	of At-Risk	Rate of	Range of	All test results		Adverse test results	
Market	Individuals	Testing	Testing	Females	Males	Females	Males
				%	%	%	%
Large	Normal	0.014	20–40	0.00129	0.00079	0.00126	0.00077
	Half	0.014	20–40	0.00251	0.00153	0.00144	0.00088
	Nil	0.014	20–40	0.00607	0.00387	0.00225	0.00142
Small	Nil	0.014	20–40	0.00547	0.00356	0.00256	0.00167
Large	Normal	0.014	20–60	0.00130	0.00080	0.00127	0.00077
	Half	0.014	20–60	0.00252	0.00154	0.00146	0.00089
	Nil	0.014	20–60	0.00623	0.00400	0.00231	0.00147
Small	Nil	0.014	20–60	0.00600	0.00367	0.00263	0.00172
Large	Normal	0.035	20–40	0.00298	0.00182	0.00283	0.00173
	Half	0.035	20–40	0.00576	0.00352	0.00319	0.00195
	Nil	0.035	20–40	0.01352	0.00861	0.00498	0.00314
Small	Nil	0.035	20–40	0.01223	0.00795	0.00574	0.00372

Table 2.18: Percentage increases in life insurance premium rates arising from severe adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. Life insurance market operating between ages 20 and 60.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Males %	Females %	Males %	Females %
Large	Normal	0.035	20–40	0.00750	0.00457	0.00723	0.00441
	Half	0.035	20–40	0.01096	0.00668	0.00819	0.00500
	Nil	0.035	20–40	0.02045	0.01293	0.01191	0.00747
Small	Nil	0.035	20–40	0.06208	0.03963	0.05558	0.03540

2.8.3 Moratoria on Using Genetic Test Results and Family History

As discussed in Section 2.5.4, the premium rates increase for two reasons when a moratorium on using genetic test results and family history is imposed. Table 2.19 shows the percentage increases in standard life insurance premium rates arising from defining new underwriting classes, and moderate and severe adverse selection, following a moratorium on the use of genetic test results and family history. The conclusions are similar to those in the case of CI insurance: the only noticeable increase results from severe adverse selection in a small market with high genetic testing rates.

Table 2.19: Percentage increases in standard premium rates for life insurance arising from new underwriting classes, and in all premiums arising from moderate or severe adverse selection, following a moratorium on the use of all genetic test results and family history. Life insurance market operating between ages 20 and 60.

	OR Premium Increase Arising From New		Premium Increase Arising From		Premium Increase Arising From	
Size of	Underwriting Classes		Moderate Adverse Selection		Severe Adverse Selection	
Market	Females	Males	Females	Males	Females	Males
	%	%	%	%	%	%
Large	0.07209	0.043098	0.02183	0.01320	0.04884	0.02954
Small	0.05649	0.034033	0.02730	0.01662	0.23896	0.14577

Chapter 3

Genetic Disorders

In Section 1.2.5, we discussed two methods of exploring the impact of genetic information on insurance, the first called "top-down", and the second called "bottom-up". The second approach requires detailed individual studies of individual genetic disorders before these studies are aggregated to assess the overall impact.

It is clearly impossible to model all genetic disorders and then aggregate them. In Section 1.2.4, we mentioned that the ABI listed eight genetic disorders regarded as significant for insurance, including early-onset familial Alzheimer's disease (EOAD), familial adenomatous polyposis (FAP), hereditary breast and ovarian cancer (BC&OC), hereditary motor and sensory neuropathy (HMSN), Huntington's disease (HD), Multiple endocrine neoplasia (MEN) and Myotonic dystrophy (MD). Starting with the ABI list, with some adjustments (see Section 1.2.5), we choose six genetic disorders, taken to have significant impact on insurance. Of these genetic disorders, four are diseases with no cause except mutations in a single gene, namely EOAD, APKD, HD and MD, and two of them are single gene subsets of common disorders, namely HNPCC and BC & OC. Our individual study of MD in Chapter 2 exemplified the methodology to be used in the "bottom-up" approach, summarized as follows:

(a) Two types of insurance policy are studied separately, CI insurance and life insurance and their corresponding insurance markets. Within each market, we ask two questions: what is the cost of insuring mutation carriers, and what is the potential cost of adverse selection, as addressed in Section 1.2.3.

(b) To answer these questions, we are interested in several aspects of the epidemiology of each genetic disorder. These aspects, as shown in Chapter 2, are the onset rates of the relevant diseases for mutation carriers, mortality rates after the onset of the relevant diseases and prevalence rate of causal mutations.

(c) The insurability of mutation carriers is addressed by modelling the life histories underlying insurance products using a Markov or semi-Markov model. The models follow the same pattern of the model shown in Figure 2.3, subject to suitable adjustment. An example of such adjustment, in the case of MD, was introduced in an extended model as shown in Figure 2.4, because the claim payment for MD is delayed until some time after onset. Please refer to other individual studies for more details.

(d) The cost of adverse selection is addressed by modeling the whole insurance market using a Markov or semi-Markov model. This model includes several factors, e.g. insurance applicants' insurance-purchasing behavior, genetic testing, development of a family history, etc. The CI market model shown in Figure 2.6 is an example.

In what follows, we aim to estimate the overall impact of genetic information using the "bottom-up" approach, and, insofar as we are able, to complete the program launched by the commencement of studies of individual genetic disorders. In this chapter, we review the epidemiology of our selected genetic disorders briefly, to add to the work on MD of the last chapter. In Chapter 4, we will bring these together to study the overall impact of genetic information.

3.1 Adult Polycystic Kidney Disease

3.1.1 Introduction to Adult Polycystic Kidney Disease

Adult polycystic kidney disease (APKD) is one of the most common single-gene dominant genetic disorders. It is also called autosomal dominant polycystic kidney disease, in contrast to the other type of polycystic kidney disease, autosomal recessive polycystic kidney disease, which is quite rare, occurs early in life and is more lethal than the dominant type.

APKD is characterized by large cysts in one or both kidneys and a gradual loss of normal kidney tissue which can lead to end-stage renal disease (ESRD, meaning kidney failure) at relatively young ages. ESRD is not treatable unless renal replacement therapy (RRT, meaning dialysis and/or a kidney transplant) is available. Two genes, each causing APKD, have been identified: APKD1 and APKD2. Mutations in APKD1 are more common, accounting for 85% of APKD, and are associated with earlier progression to ESRD. It is possible that another mutation, APKD3, exists, which might also cause APKD, but so far there is no concrete evidence. An accurate diagnosis of APKD relies on ultrasonic imaging or molecular genetic testing. The accuracy of ultrasonic testing is nearly 100% for all patients with APKD aged above 30 years (Gutiérrez & Macdonald, 2007).

3.1.2 Epidemiology of APKD

Estimates of Onset Rates Associated with APKD1 and APKD2 mutations

Gutiérrez & Macdonald (2003) used studies of the onset rate of APKD that did not differentiate between mutations in the APKD1 and APKD2 genes. This paper only considered the CI insurance market. Two data sources were used: Churchill *et al.* (1984) and the

U.S. renal diseases system (USRDS).

Gutiérrez & Macdonald (2007) improved the results in Gutiérrez & Macdonald (2003) by using studies that distinguish between mutations in the two different genes. In this paper, the sources were Hateboer *et al.* (1999), Johnson & Gabow (1997) and Ravine *et al.* (1992). In view of the very small samples of APKD2 subjects in Johnson & Gabow (1997) and Ravine *et al.* (1992), Hateboer *et al.* (1999) was taken to be the most reliable study of APKD2. We also will use only Hateboer *et al.* (1999) as a source for APKD1. Figure 3.1 (top) shows a curve fitted to Kaplan-Meier estimates of age-specific survival probabilities where the 'event' is the earlier of ESRD or death. Therefore the intensity derived from this fitted curve, called 'quasi onset rate', is not just the onset intensity of ESRD, but the summation of the onset intensity of ESRD and the force of mortality. Therefore the force of mortality must be subtracted from the 'quasi onset rate' to find the onset rate of ESRD. The force of mortality here is that of the English Life Table No.15 for women and men in normal health, because heavier mortality is accounted for by the post-onset mortality. Figure 3.1 (bottom) shows the adjusted onset rate of ESRD for APKD1 and APKD2 mutation carriers. See Appendix C.1 for more technical details.

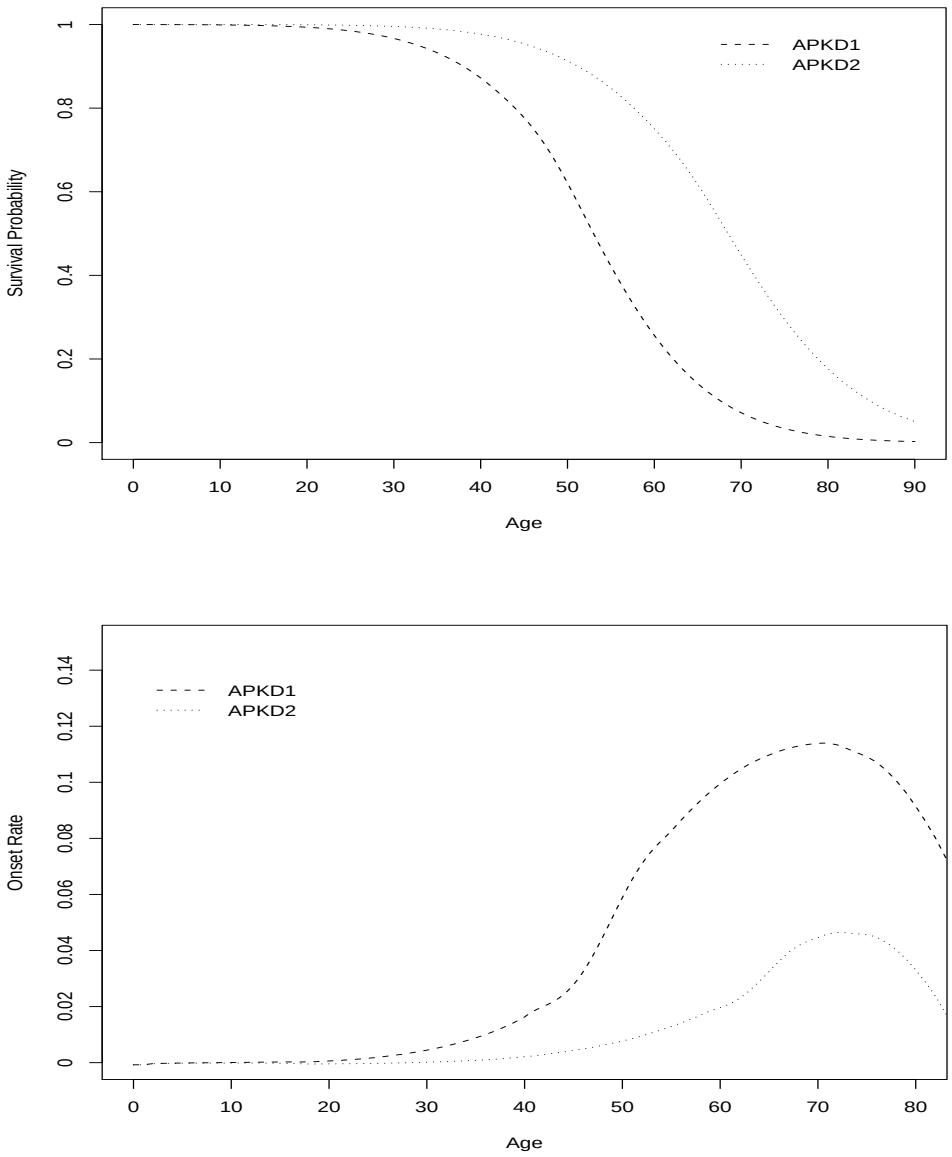

Figure 3.1: Parametric functions fitted to Kaplan-Meier estimates of the probabilities of surviving ESRD or death for APKD1 and APKD2 mutation carriers (top). Fitted onset rates of ESRD for APKD1 and APKD2 mutation carriers, adjusted using the mortality of English Life Tables No.15 (bottom). Data source: Hateboer *et al.* (1999)

Post-Onset Mortality Rate

The post-onset mortality rates depend on the treatment after ESRD and also whether it is short-term or long-term (Gutiérrez & Macdonald, 2007). APKD patients have two options after ESRD, dialysis or renal transplant (RRT).

In the short-term study by Tsakiris *et al.* (1999), 1.4% of the sampled patients died within 90 days. In a long-term study, Canadian Organ Replacement Register (CORR) carried out Kaplan-Meier estimation for patients receiving either dialysis or renal transplant treatment. These Kaplan-Meier estimates were fitted using Gamma functions in Gutiérrez & Macdonald (2007). Be aware that these estimations are conditional on patients having survived 90 days after the treatment. The post-onset mortality rates could be derived directly based on these estimates. One important feature is that these survival rates are duration-dependent. For brevity, we call these rates $\mu_{x,z}^{APKD,Mortality}$, where z denotes the duration after the onset of ESRD. Please refer to Appendix C.1 for more technical details. Although these rates could be applied directly, Wilkie (2000) pointed out the anomaly that $\mu_{x,z}^{APKD,Mortality}$ could sometimes be substantially lower than the normal age-related mortality rates $\mu_x^{Standard}$ (e.g. English Life Tables 15) at certain ages. Therefore, to avoid this anomaly, we assume that mortality after onset of ESRD is no better than normal age-related population mortality, i.e. $\mu_{x,z}^{APKD,Death} = \max\left\{\mu_{x,z}^{APKD,Mortality}, \mu_x^{Standard}\right\}$. Figure 3.2 demonstrates the Gamma functions fitted to Kaplan-Meier estimates of survival probability functions for ESRD patients receiving either dialysis or RRT treatment. Figure 3.3 shows the derived force of mortality $\mu_{x,z}^{APKD,Mortality}$ based on Figure 3.2.

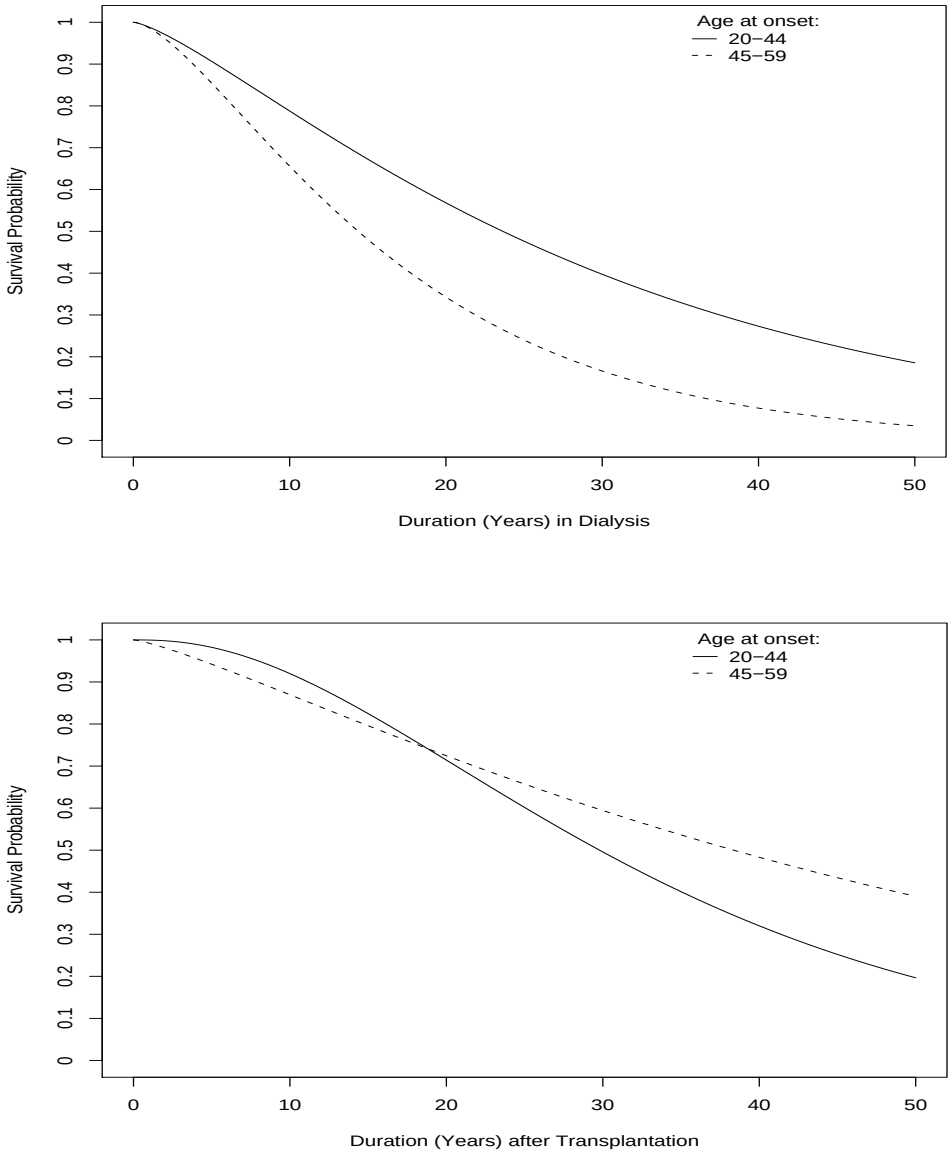

Figure 3.2: Gamma functions fitted to Kaplan Meier estimates of survival probability function for ESRD patients receiving treatments, either dialysis or transplantation. Both cases have two age-at-onset groups, 20–44 and 45–59.

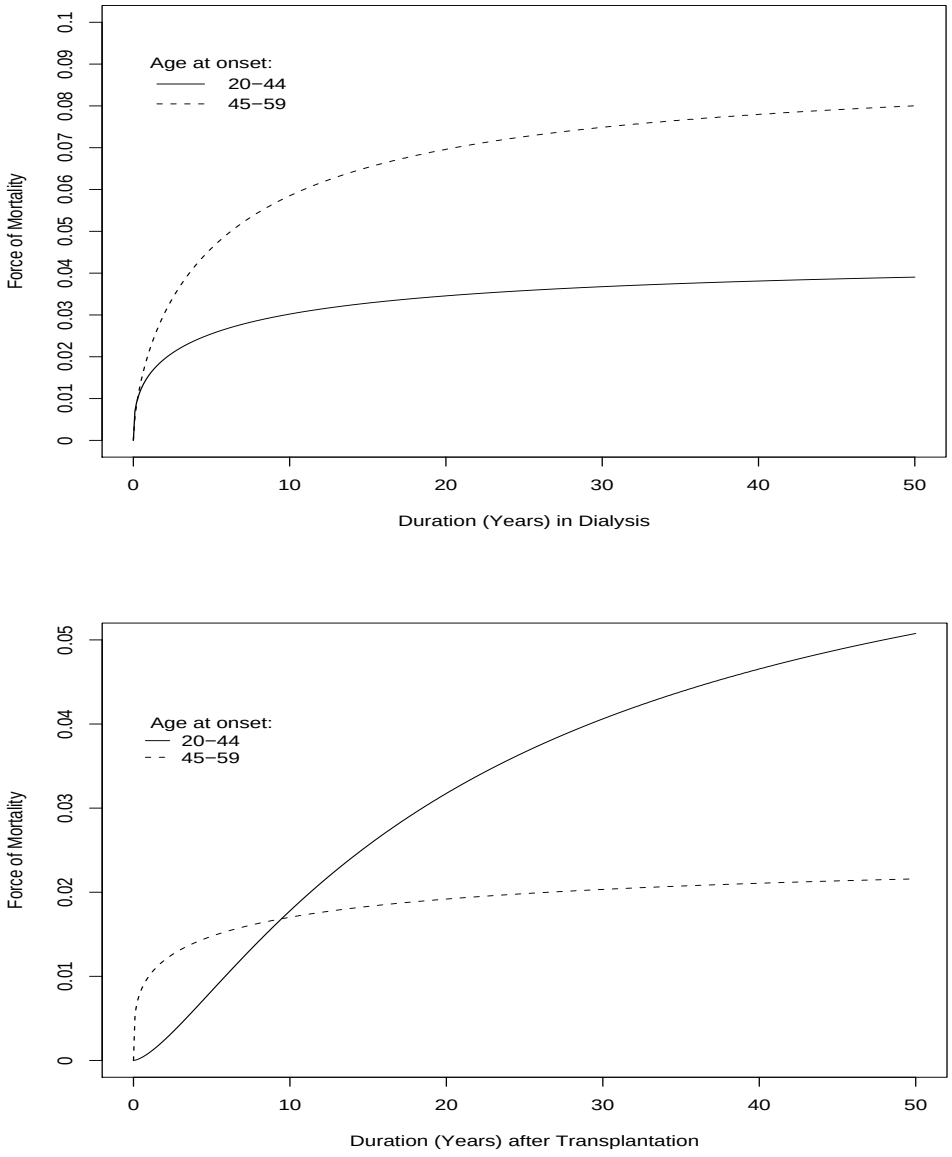

Figure 3.3: The derived force of mortality $\mu_{x,z}^{APKD,Mortality}$ based on Figure 3.2 for ESRD patients receiving treatment, either dialysis or transplantation. Both cases have two age-at-onset groups, 20–44 and 45–59.

For the intensity from ESRD to RRT, Gutiérrez & Macdonald (2007) assumed four scenarios, because this transition would depend on the availability of kidneys for transplantations, which varies from place to place.

(a) The intensity is 0.

(b) The intensity is low and is about 0.05 per year.

(c) The intensity is high and is about 0.15 per year.

(d) The intensity is extremely high and all the APKD patients can have renal transplant directly upon ESRD.

For simplicity, in this thesis, we use only scenario (b), i.e. the intensity of transferring from ESRD to RRT is 0.05 per year.

Prevalence Rate and the Distribution of Mutations

Gutiérrez & Macdonald (2003) and Gutiérrez & Macdonald (2007) both use the same aggregate prevalence rate, 1 per 1,000 of the population (Dalgaard, 1957). APKD1 mutations are both more common than APKD2 mutations (they account for about 85% and 15% of APKD, respectively) as well as being more severe. Therefore we take the prevalence rates for APKD1 and APKD2 mutations to be 0.00085 and 0.00015, respectively.

3.2 Early-onset Alzheimer's Disease

3.2.1 Introduction to EOAD

Alzheimer's disease (AD) is named after Dr. Alois Alzheimer. The most striking early symptom is the loss of memory (dementia), accompanied by other symptoms, such as disorientation and it will progress to the domain of language, skilled movements, recognition, and those functions closely related to the frontal and temporal lobes of the brain. The Alzheimer's Society (`http://www.alzheimer's.org.uk`) estimates that over 700,000 people in the UK have dementia. AD is a complicated disorder. Normally, it can be further classified into late- and early-onset forms, depending on whether the onset age is after or before 65 years old.

For late-onset AD (LOAD), the main genetic factors known so far are connected with the Apolipoprotein E (ApoE) gene, which has three common alleles — $\varepsilon 2$, $\varepsilon 3$ and $\varepsilon 4$, of which $\varepsilon 4$ is the most important genetic risk factor for AD. Since each offspring receives one allele from each parent, there are six possible genotypes $\varepsilon 2/\varepsilon 2$, $\varepsilon 2/\varepsilon 3$, $\varepsilon 2/\varepsilon 4$,$\varepsilon 3/\varepsilon 3$, $\varepsilon 3/\varepsilon 4$ and $\varepsilon 4/\varepsilon 4$. Each of these combinations of mutations carries a different risk of LOAD. See Macdonald & Pritchard (2000) for more details.

For early-onset AD (EOAD), mutations in three genes have been reported, namely the amyloid Precursor Protein (APP), Presenilin-1 (PSEN-1) and Presenilin -2 (PSEN-2) genes. Mutations in these genes lead to increased production of a protein fragment called amyloid beta peptide, which is found in the brain and other tissues. This peptide can build up in the brain to form clumps called amyloid plaques. Over 130 mutations have been described in PSEN-1, only 10 in the PSEN-2 gene and about 20 in APP (Armstrong *et al.*, 2004, Giacomello *et al.*, 2005 and Espinosa-Castañeda, 2006). PSEN-1 is implicated in 18–55% of all families with EOAD, PSEN-2 is very rare, and APP mutations account for 10–15% of all families with EOAD (Bird, 2005 and Espinosa-Castañeda, 2006).

Diagnosis is made primarily on the basis of family history, clinical observation, memory tests etc. Experts who specialize in memory disorders can now diagnose AD with an accuracy of 85–90%, but a definitive diagnosis of AD must use microscopic examination of brain tissue, only permitted post-mortem.

Patients with EOAD normally live 5 to 10 years from the time of diagnosis (Pasternak, 1999). Death often follows from cardiovascular disease, bronchitis and pneumonia. Death may also be attributable to the terminal complications of AD, consisting of a progressive vegetative state, malnutrition and dehydration, culminating in cardiorespiratory arrest (Heyman *et al.*, 1987).

There is currently no cure for Alzheimer's disease. Most treatments can only make patients comfortable, but do not show any signs of slowing down the progression of the disease, in particular, alleviating the level of dementia. Physical exercise, intellectual stimulation (e.g. chess or crosswords, regular social interaction) are proven to be effective ways of reducing the risks of AD.

3.2.2 Epidemiology of EOAD

Onset Rates

In Gui & Macdonald (2002), a large number of pedigrees was collected from published papers. The authors estimated the integrated onset rate using Nelson-Aalen methods based on these pedigrees, and hence the onset rate itself. Two items of information, in respect of each life sampled, were needed:

(a) The age at which each person entered the risk group, which is their age when their parent or first sibling contracted EOAD.

(b) The age at onset of EOAD, or the age at censoring in other cases.

Then the exposure time can be calculated according the above information. The problem is that neither of these two items is straightforward. In most cases, the only information the authors have is pedigrees, which do not specify this information. Therefore both items requires approximation if the information is not given or cannot be approximated in a better way, hence the exposure time has to be approximated. In this situation, the exposure time is not an accurate single figure, but a range, with its minimum and maximum values, called minimum exposure times and maximum exposure times, respectively. The authors aimed to calculate the integrated onset rate based on these extremes and then define a feasible region for the onset rate.

In Gui & Macdonald (2002), the estimation was based on certain unrealistic assumptions, in particular no ascertainment bias and full penetrance. Therefore the probability of the an at-risk person carrying a mutation, denoted p, was assumed equal to $1/2$, and the lifetime penetrance, denoted σ, was assumed to be 1. Espinosa & Macdonald (2007) used the same data and calculated bias-corrected estimates of onset rates with different assumptions of lifetime penetrance σ and probability p. It is more realistic to assume that $p > 1/2$ and $\sigma < 1$ because of the retrospective methods of ascertainment used to select the pedigrees. The onset rates were estimated with four different choices of σ in Espinosa & Macdonald (2007), (1) $\sigma = 100\%$, (2) $\sigma = 90\%$, (3) $\sigma = 80\%$ and (4) $\sigma = 0.653$ for maximum exposures, $\sigma = 0.728$ for minimum exposures. Espinosa & Macdonald (2007) showed that in case (1) the estimates are too high because of unrealistic assumptions, and that in cases (2), (3) and (4) the onset rates are not hugely different. In this thesis, for simplicity, we use only a lifetime penetrance of 80%. Please refer to Espinosa & Macdonald (2007) for more technical details. Figure 3.4 shows the integrated intensities and smoothed versions of the estimator, in which (a) is the minimum exposure case, and (b), the maximum exposure case. Figure 3.5 shows the survival probabilities, assuming the lifetime penetrance is 0.8; again (a) shows the minimum exposure case, and (b), the maximum exposure case. Figure 3.6 shows the corresponding onset rates of EOAD.

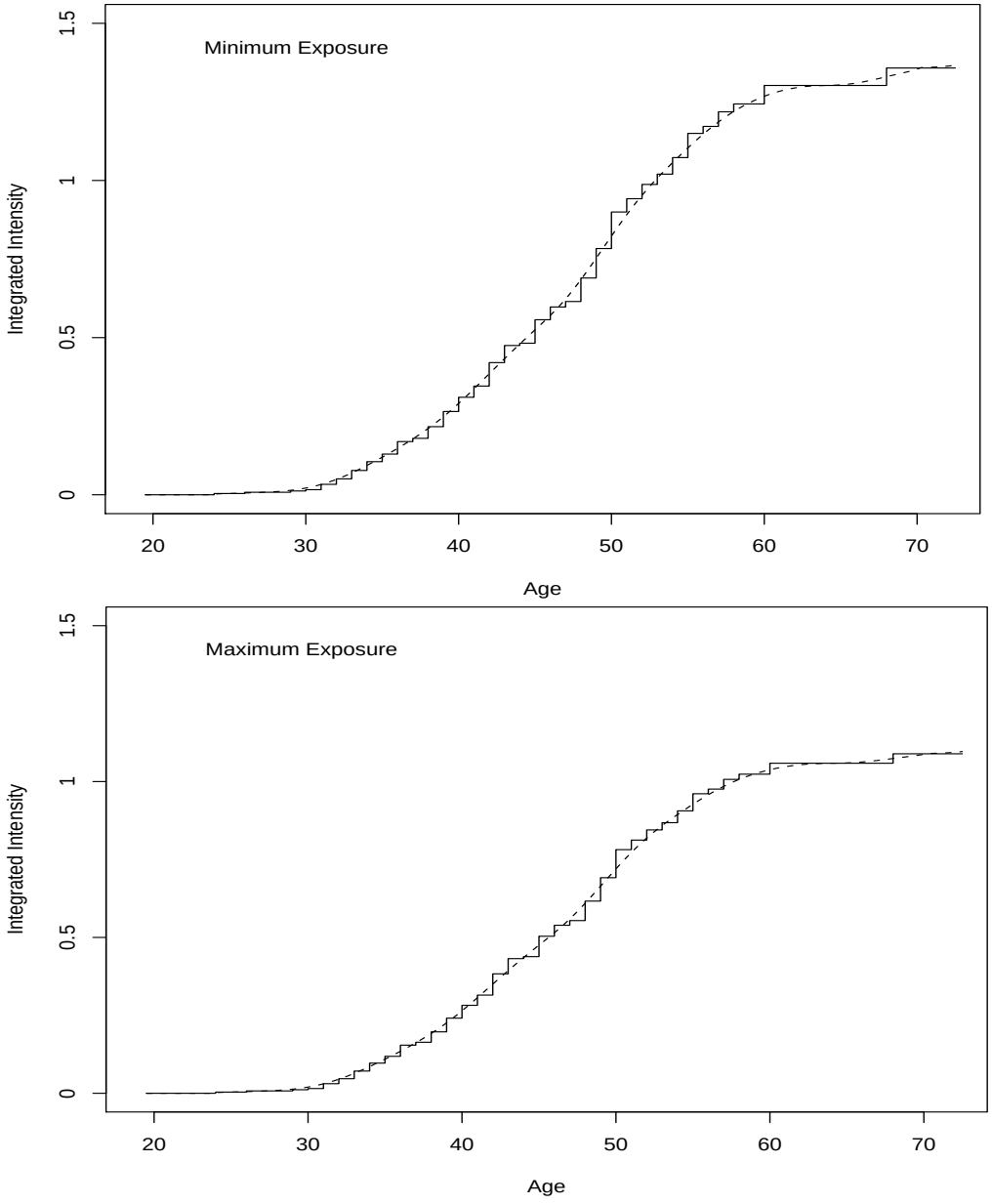

Figure 3.4: Modified Nelson-Aalen Estimates of the integrated intensity of onset of EOAD, in respect of PSEN-1 mutation carriers, with minimum (top) and maximum (bottom) exposures. The smoothed versions of the estimator are shown as dotted lines.

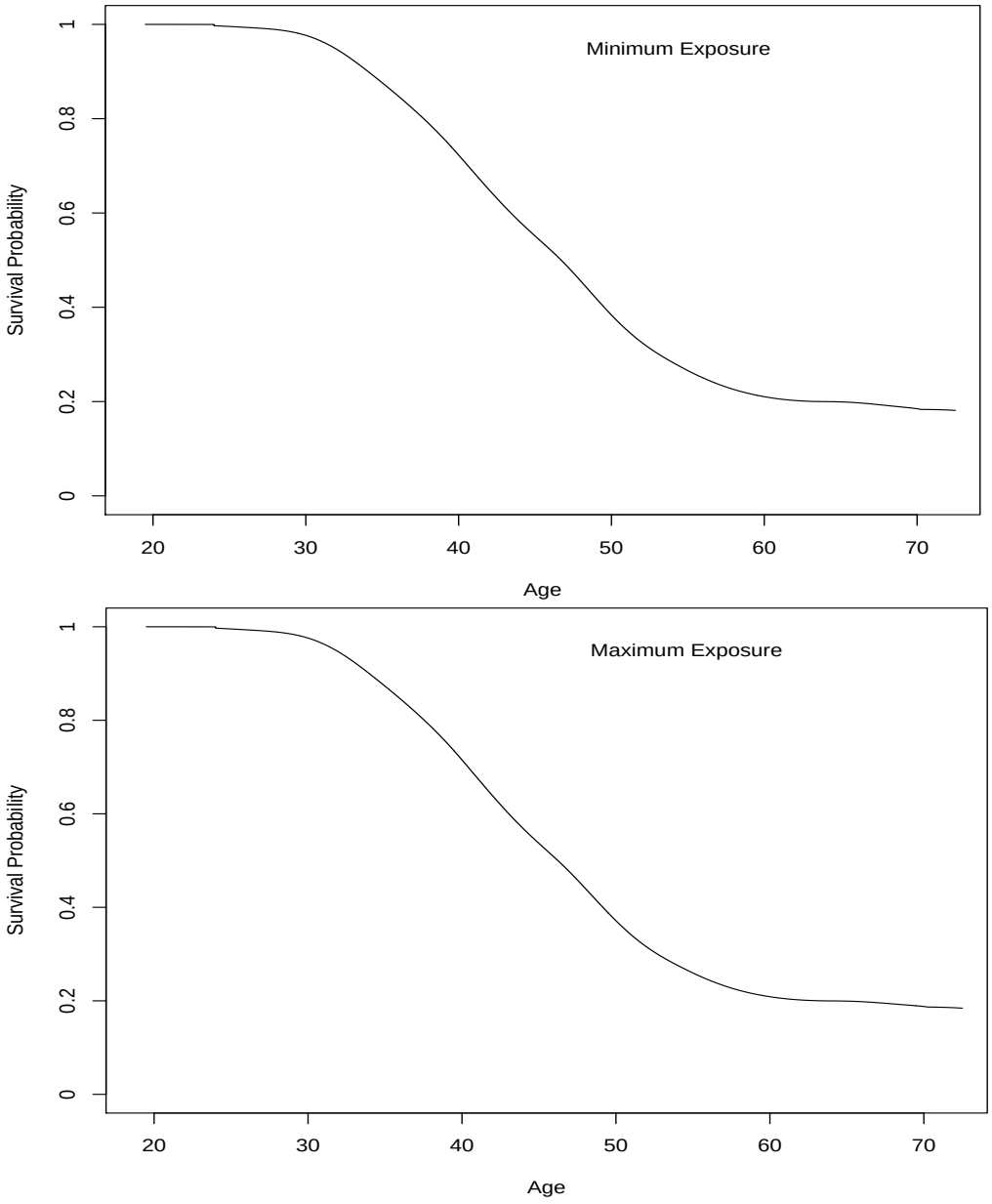

Figure 3.5: Survival functions for onset of EOAD, in respect of PSEN-1 mutation carriers, with minimum (top) and maximum (bottom) exposures. The assumed lifetime penetrance is 0.8.

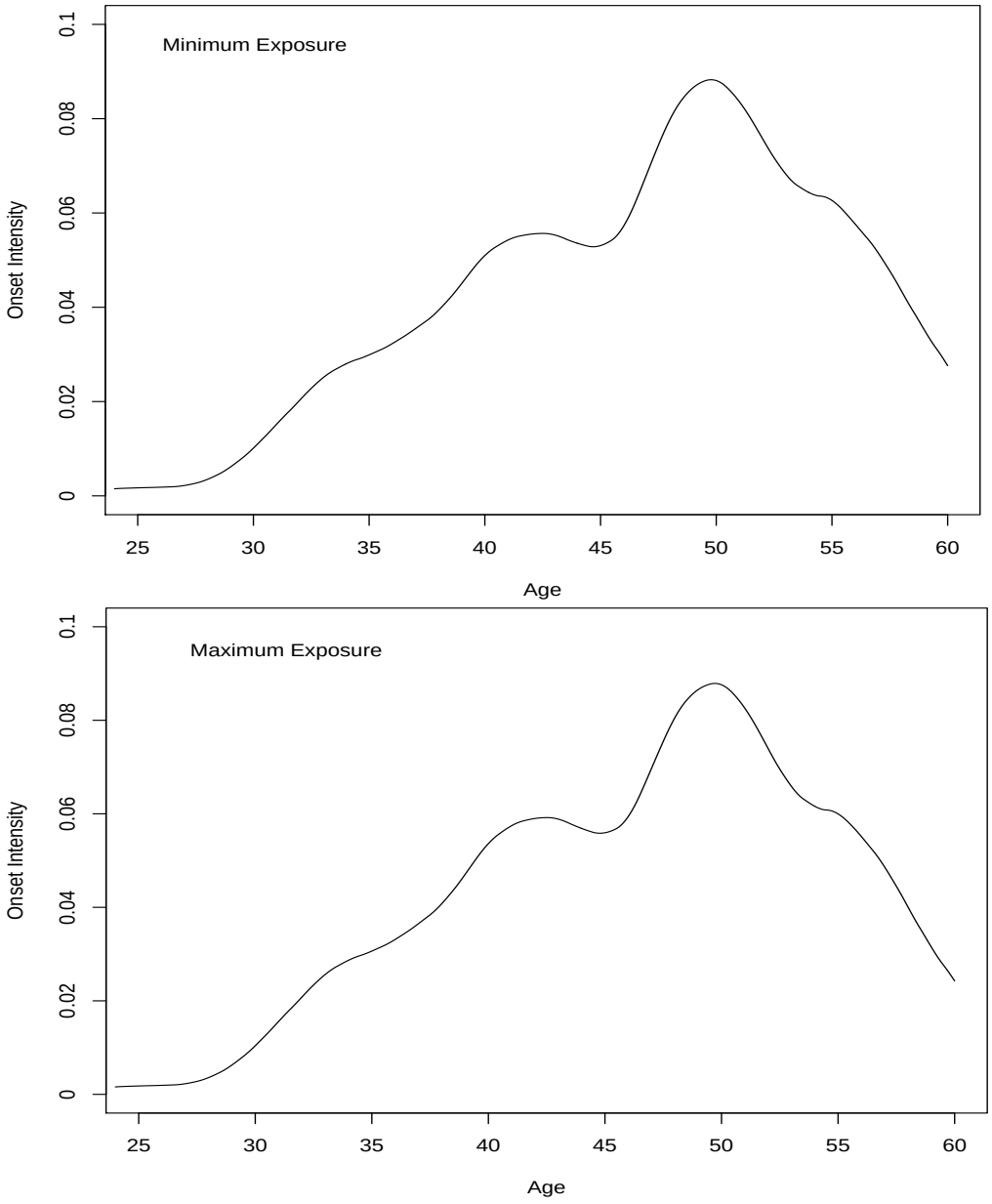

Figure 3.6: Estimated onset intensities of EOAD, in respect of PSEN-1 mutation carriers, with minimum (top) and maximum (bottom) exposures. The assumed lifetime penetrance is 0.8.

Post-Onset Mortality

In Gui & Macdonald (2002b), a duration-dependent mortality rate after onset is derived. We call it $\mu_z^{EOAD,Mortality}$ and it is shown in Figure 3.7, where z denotes the duration after onset of EOAD. Please refer to Appendix C.2 for more technical details. As in the case of APKD and MD, post-onset mortality $\mu_z^{EOAD,Mortality}$ should not be assumed to be lower than normal population mortality, where the population mortality is that of ELT15. Therefore we take $\mu_{x,z}^{EOAD,DEAD} = \max\left\{\mu_z^{EOAD,Mortality}, \mu_x^{Standard}\right\}$.

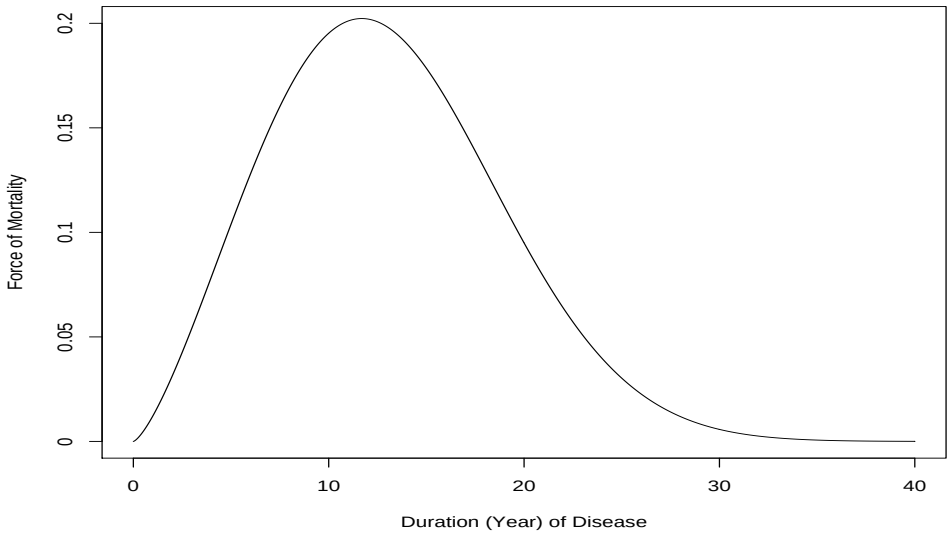

Figure 3.7: Force of mortality $\mu_z^{EOAD,Mortality}$ for PSEN-1 gene mutation carriers after onset of EOAD, as a function of duration z since onset.

Prevalence Rate

Gui (2003) estimated the prevalence rate of PSEN-1 mutation carriers based a population-based prevalence study by Campion *et al.* (1996). This study estimated the prevalence

of familial EOAD at 25.4 per 100,000 at risk. Of the 184 affected subjects tested for mutations by Campion *et al.* (1996), 14.7% carried APP gene mutations while 58.8% had PSEN-1 gene mutations. Hence the prevalence rate of PSEN-1 mutation carriers was estimated at 15 per 100,000. Espinosa-Castañeda (2006) also used 15 per 100,000 as the prevalence rate of PSEN-1 mutation carriers.

3.3 Huntington's Disease

3.3.1 Introduction to Huntington's Disease

Huntington's disease (HD), also called Huntington disease, or Huntington chorea, takes its name from the New York physician George Huntington who described it precisely in 1872 in his first medical paper. This disease has been heavily studied during the past decades. HD's most obvious symptoms are abnormal body movements called chorea, and a lack of coordination, but it also affects a number of mental abilities and some aspects of personality. This disease can occur at any age. HD patients do not lose their control of movement suddenly, but endure the long-term progression of the disease. The normal symptom of onset is the uncontrolled body movement, and later symptoms affect the cognitive system, mainly executive functions (e.g. planning, abstract thinking). Psychopathological symptoms will appear, such as reduced display of emotions, anxiousness and depression (Harper, 1996).

HD is a typical single-gene autosomal dominant genetic disorder. It is caused by mutations in the Huntingtin gene. In a certain region of the gene, there is a sequence consisting of a variable number of CAG trinucleotides. Normally the gene has fewer than 35 CAG repeats, but when this number exceeds 35, people may develop HD. When the number of CAG repeats is 36–39, the penetrance is incomplete, and when the number of CAG repeats is higher than 40, the onset of HD is practically certain, that is, the mutation is completely penetrant. Therefore different CAG repeat lengths lead to different onset rates.

Diagnosis of this disease normally begins with the recognition of unusual body movements. But when the disease progresses to affect cognitive functions, and further psychopathological functions, diagnosis is not easy, because of the potential confusion with other neurological diseases. Genetic testing now provides the most reliable diagnosis. By

counting the number of repetitions of the CAG triplet in the Huntingtin gene, we can show conclusively whether the person tested is a mutation carrier. As to the prognosis of HD, the onset age is correlated with the number of CAG repeats. The longer the CAG repeats expansion, the earlier onset is likely to be.

HD does not have any effective treatment. So far no treatment can reduce the risks of being affected by HD, but some may alleviate the symptoms, and make the patients more comfortable. Maintaining high quality nutrition is an important part of treatment. Other things, e.g. speech training, also help patients to maintain their speaking ability. The long-term potential treatment may lie in genetic therapy. But so far these methods exist only theoretically, and are still in the process of being tested on animal models.

3.3.2 Epidemiology of HD

Onset Rate

Brinkman *et al.* (1997) used Kaplan-Meier methods to estimate the cumulative probability of surviving without HD, categorized by numbers of CAG repeats. In Gutiérrez & Macdonald (2004), the authors fitted the penetrance with Gamma distributions, and then derived the onset rates of HD for different numbers of CAG repeats, because they were interested in the effect of genetic information relating to HD only. Alternatively, we could estimate an aggregate onset rate. In this thesis, we are interested in the overall effect of genetic information relating to not only HD, but all chosen genetic disorders, so that we choose to do the latter, and derive the aggregate onset intensity from the aggregate age-at-onset distribution. Several large studies have found that the age at onset distribution is very close to Normal, e.g. Bell (1934), Wendt & Drohm (1972). The aggregate age-at-onset distribution we will use is that is parameterized by MacCalman (2009). Please see Appendix C.3 for more technical details. Figure 3.8 shows the age-at-onset distribution (top) and the derived onset rate of HD for mutation carriers (bottom).

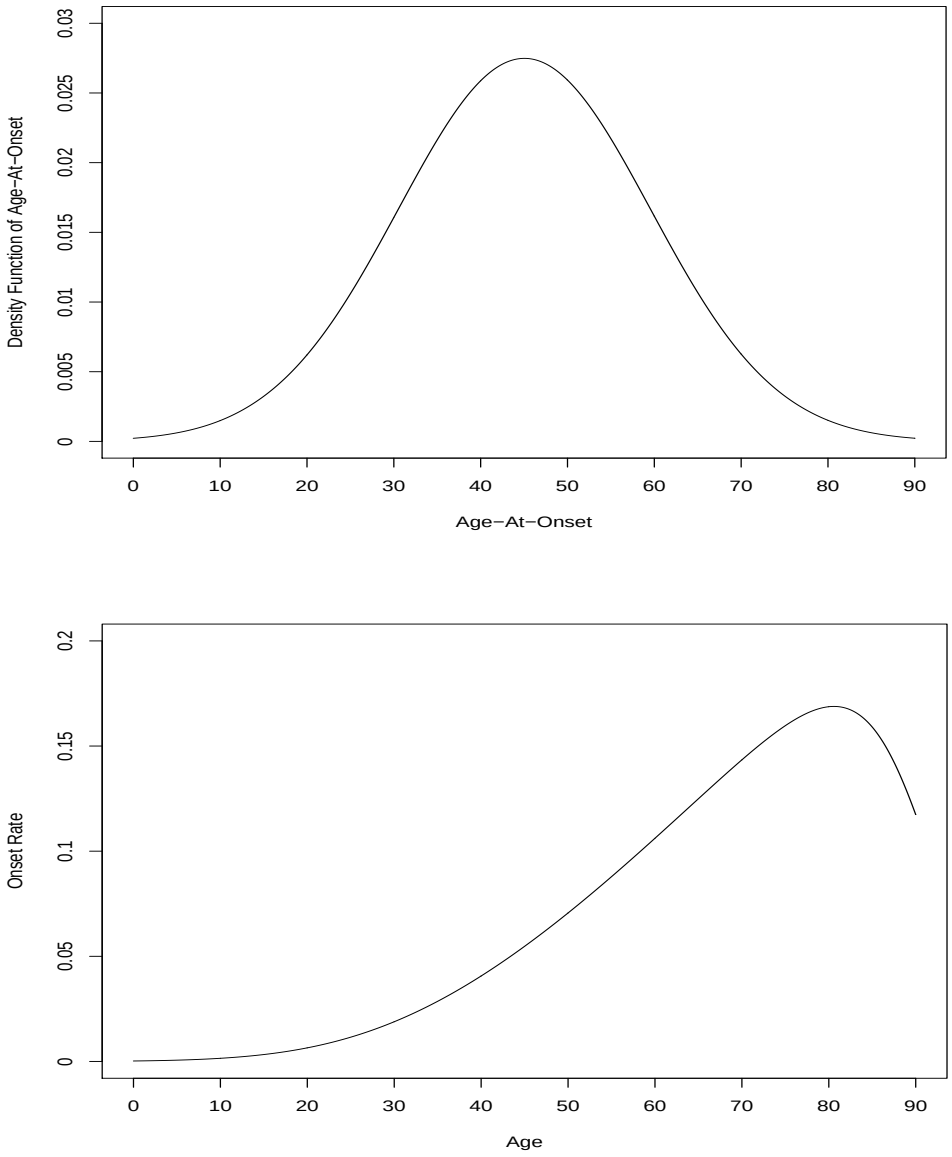

Figure 3.8: Normal age-at-onset distribution (top) and associated onset rate of HD (bottom) for HD mutation carriers. Source: MacCalman (2009).

Post-Onset Mortality

The post-onset mortality rate $\mu_{x,z}^{HD,Mortality}$ is estimated in Gutiérrez & Macdonald (2004), based on Foroud *et al.* (1999), which estimated the survival probabilities as a function of duration since onset of HD for three age-at-onset groups, 20–34, 35–49 and above 50. The latter study was chosen because it used a definition of "onset" of HD matching that used in the studies of the onset of HD. Please see Appendix C.3 for more technical details. Figure 3.9 (top) shows the fitted survival probability for three age-at-onset groups, and Figure 3.9 (bottom) displays the post-onset mortality rates $\mu_{x,z}^{HD,Mortality}$, or force of mortality, for these three groups. As in other cases, post-onset mortality $\mu_{x,z}^{HD,Mortality}$ should not be assumed to be lower than normal population mortality, where the population mortality is that of ELT15. Therefore we take $\mu_{x,z}^{HD,DEAD} = \max\left\{\mu_{x,z}^{HD,Mortality}, \mu_x^{Standard}\right\}$.

Prevalence Rate

Gutiérrez & Macdonald (2004) cited an overall mutation prevalence rate of 18.75 per 100,000 population (Harper, Lim & Craufurd, 2000). This figure is based on a disease prevalence of about 7.5 per 100,000, where disease prevalence means the proportion of people with symptoms of HD at any fixed calendar time. Since others, especially at younger ages, will carry a HD mutation but not yet show symptoms, the mutation prevalence in the population is much higher. Harper (1996) suggests about 2.5 times higher, hence the HD mutation prevalence of 18.75 per 100,000.

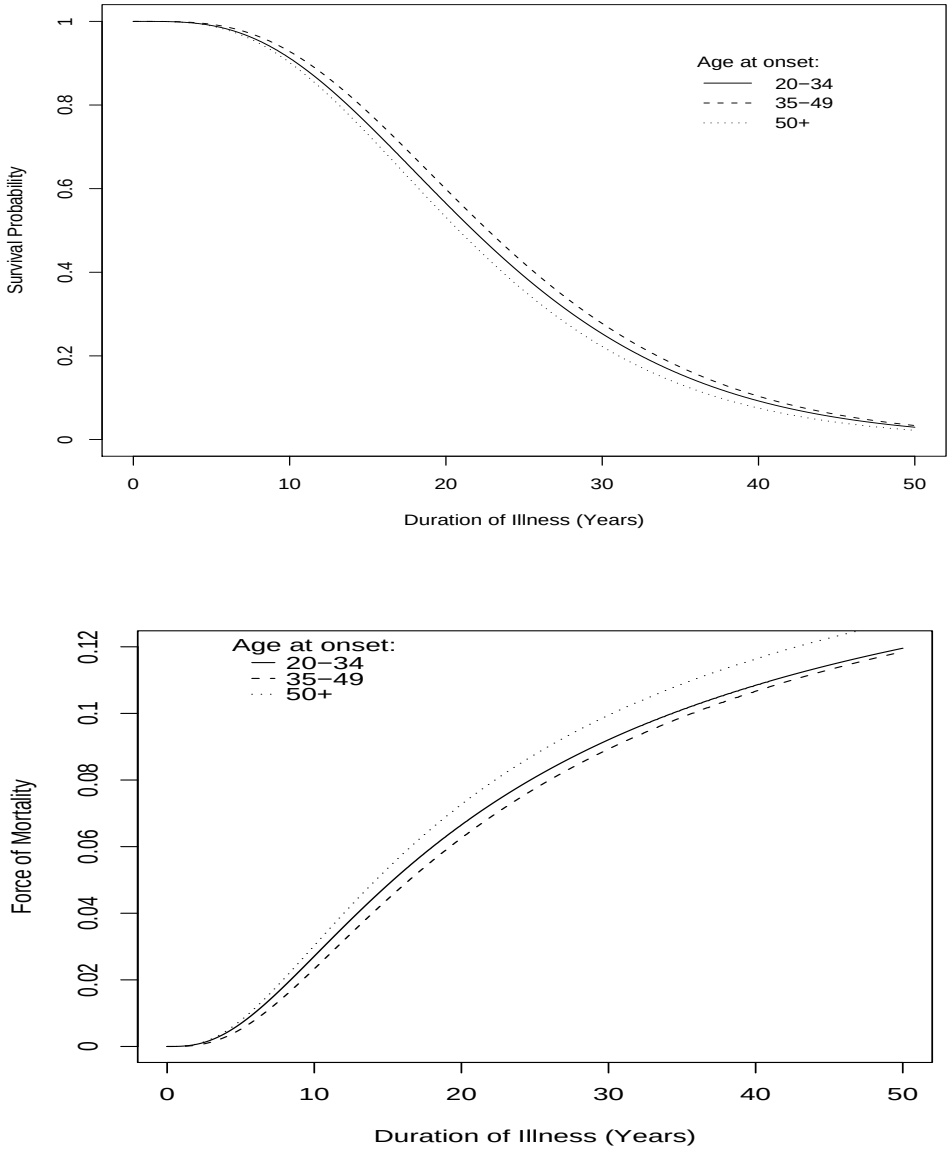

Figure 3.9: The survival probabilities as a function of duration since onset of HD for three age-at-onset groups (top) and corresponding post-onset mortality $\mu_{x,z}^{HD,Mortality}$ (bottom), based on Foroud *et al.* (1999)

3.4 Hereditary Non-Polyposis Colorectal Cancer

3.4.1 Introduction to HNPCC

Colorectal cancer (CRC) is cancer of the colon or rectum. CRC is the second commonest cancer affecting women (after breast cancer) and the third commonest cancer in men (after prostate and lung cancer). Colorectal cancers can be sporadic or hereditary. Approximately 6% of colorectal cancers are hereditary, of which around 83% are accounted for by Hereditary Non-Polyposis Colorectal cancer (HNPCC), and the rest by Familial Adenomatous Polyposis (FAP) (Lynch & Smyrk, 1996a).

FAP is an inherited condition in which numerous polyps grow in the colon. While these polyps start out benign, malignant transformation into colon cancer occurs in 100% of cases if not treated. FAP can have different inheritance patterns and genetic causes. When this condition results from mutations in the APC gene, it is inherited in an autosomal dominant pattern, which means one copy of the altered gene is sufficient to cause the disorder. Colonoscopy is considered the diagnostic test of choice as it can provide not only a quantification of polyps throughout the colon but also a histological diagnosis. Most individuals with the APC mutation will develop colon cancer by the age of 40. Therefore, prophylatic surgery is generally recommended before the age of 25 (Gebert & Knebel Doeberitz, 1999). In this thesis, we do not model FAP, because these polyps can be easily detected and removed in a screening program, and the risk associated with FAP is limited (see Section 1.2.5 for details).

HNPCC may be caused by mutations in DNA mismatch repair (MMR) genes or by non-genetic factors, such as life style. Mutations in five MMR genes, namely MSH2, MLH1, PMS1, PMS2 and MSH6 have so far been identified to be responsible for most cases of HNPCC, with MLH1 and MSH2 mutations accounting for 90% of cases (Lynch & de la Chapelle, 2003). One important feature of HNPCC is that the mutation carriers

are not only at risk of CRC, but also are prone to develop endometrial cancer (females only), brain cancer, small bowel cancer, gastric cancer, upper urinary tract cancer, and extracolonic cancers, e.g. ovarian cancer (female only).

The 'Amsterdam criteria' have been used to identify HNPCC patients in clinical practice. They have three major requirements:

(a) histologically verified CRC in three or more relatives, one of whom is first-degree relative of the other two;

(b) CRC in at least two generations; and

(c) one or more CRC cases diagnosed before age 50.

Recent developments suggest that the Amsterdam criteria are too restrictive, and underestimate the prevalence of HNPCC, because they ignore the extracolonic cancers, and small families rarely meet them (Lu *et al.*, 2007). Improved criteria have been proposed, such as the modified Amsterdam criteria (Bellacosa, Genuari & Anti, 1996) and Bethesda criteria (Rodriguez-Bigas, Boland, & Hamilton, 1997).

More advanced methods to identify HNPCC are the MSI test (Ikenaga *et al.*, 2002) and genetic testing (Hoedema *et al.*, 2003). One of the interesting features of HNPCC patients is microsatellite instability (MSI), while FAP patients normally show microsatellite stability (MSS). MSI is almost ubiquitous in tumours from patients who are members of HNPCC families. Liu, Parsons and Paradpoulos (1996) discovered that 92% of CRCs in HNPCC patients show the MSI phenotype. Farrington *et al.* (1998) found 65% of young patients with MSI tumours had germline MSH2 or MLH1 mutations. However, for sporadic CRC patients, 10–15% of cases show MSI (Lynch & Smyrk, 1996b and Papadopoulos, 2004).

As to the prognosis, several papers suggest that the survival rate of HNPCC patients is better than that of sporadic CRC patients (Sankila *et al.*, 1996, Lin *et al.*, 1998,

Watson *et al.*, 1998, Farrington *et al.*, 1998, Lynch & Smyrk, 1996a, Percesepe *et al.*, 1997, Bertario *et al.*, 1999, Tomoda, Baba & Akazawa, 1999 and Elsakov & Kurtinaitis, 2006). The possible reason is that HNPCC patients can be identified and treated earlier than patients with sporadic CRC.

'Staging' is a method to evaluate the severity of cancer in a patient. Two commonly used CRC staging systems are the Duke Staging System and the TNM Staging System. The Duke staging system originally was published by C.E. Dukes in 1932 for rectal cancer only. It was adapted by Kirklin in 1949 and later by Astler and Coller in 1953 for cancers of the colon and rectum (Astler & Coller, 1954). Their work contributed to the Modified Duke System. The TNM System was recently developed by the American Joint Committee on Cancer (AJCC) (Watson *et al.*, 1998).

Treatment depends on the staging of CRC. When colorectal cancer is detected at early stages (with little spread) it can be curable. However when it is detected at later stages (when distant metastases are present) it is less likely to be curable. Surgery remains the primary treatment while chemotherapy and/or radiotherapy may be recommended.

CRC screening is an effective way to detect CRC at its early stages and so to reduce the CRC associated risks. CRC screening programs include the Bowel Screening Program for the general population, and the HNPCC Surveillance Program for HNPCC families. Several papers claim that the people participating in a screening program will have better prognosis than those who choose not to do so (Järvinen, Mecklin & Sistonen, 1995, Järvinen *et al.*, 2000, de Vos tot Nederveen Cappel *et al.*, 2002, Vasen *et al.*, 1993 and Renkonen-Sinisalo *et al.*, 2000). In this chapter, we study the post-onset mortalities for both HNPCC patients and sporadic CRC patients, assuming the sampled persons choose not to take part in any surveillance program. In Chapter 5, we will study the effect of CRC screening programs on CRC associated risks in more detail.

3.4.2 Epidemiology of HNPCC

Onset Rates

Lu (2006) estimated the onset rate for both MLH1 and MSH2 mutation carriers and non-mutation carriers. For mutation carriers, the estimates were based on data from Vasen *et al.* (2001). People carrying MLH1 and MSH2 mutations are not only at risk of colorectal cancer (CRC), but also are more likely to contract endometrial cancer (EC) (females only) and extracolonic cancer (OECC) (including cancer of the stomach, urinary tract, small bowel, ovary (female only) and brain). See Appendix C.4 for more technical results. Figure 3.10 shows the fitted cumulative onset probability functions (top) and the derived onset rates (bottom) for both male and female MLH1 and MSH2 mutation carriers, in respect of CRC. Figure 3.11 shows the fitted cumulative onset probability functions (top) and onset rates (bottom) for female MLH1 and MSH2 mutation carriers, in respect of EC. Figure 3.12 shows the fitted cumulative onset probability functions (top) and onset rates (bottom) for both male and female MLH1 and MSH2 mutation carriers, in respect of OECC. Figure 3.13 shows the fitted population onset rates of CRC for males and females and fitted population onset rates of EC for females.

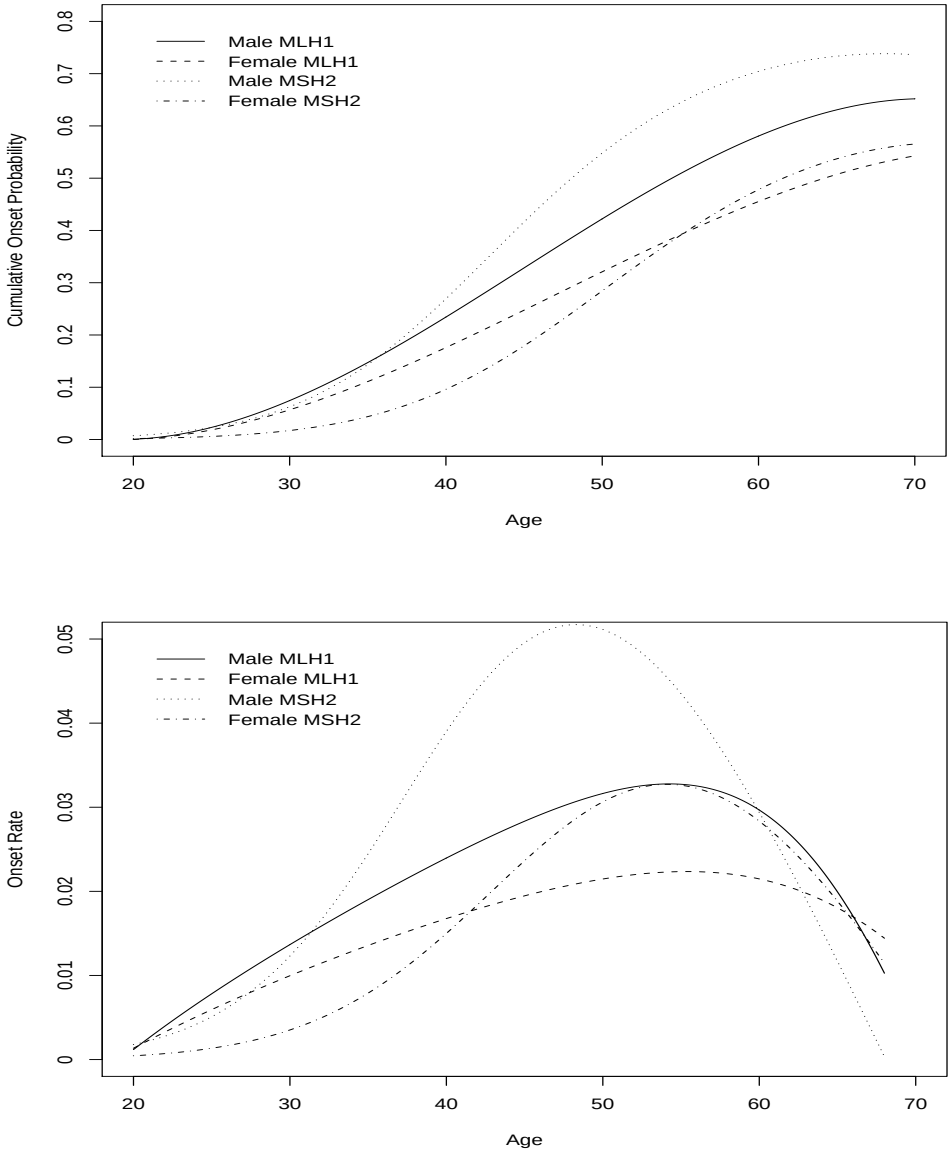

Figure 3.10: Fitted cumulative onset probability functions (top) and onset rates (bottom) for both male and female mutation MLH1 and MSH2 carriers, in respect of CRC. Source: Lu *et al.* (2007).

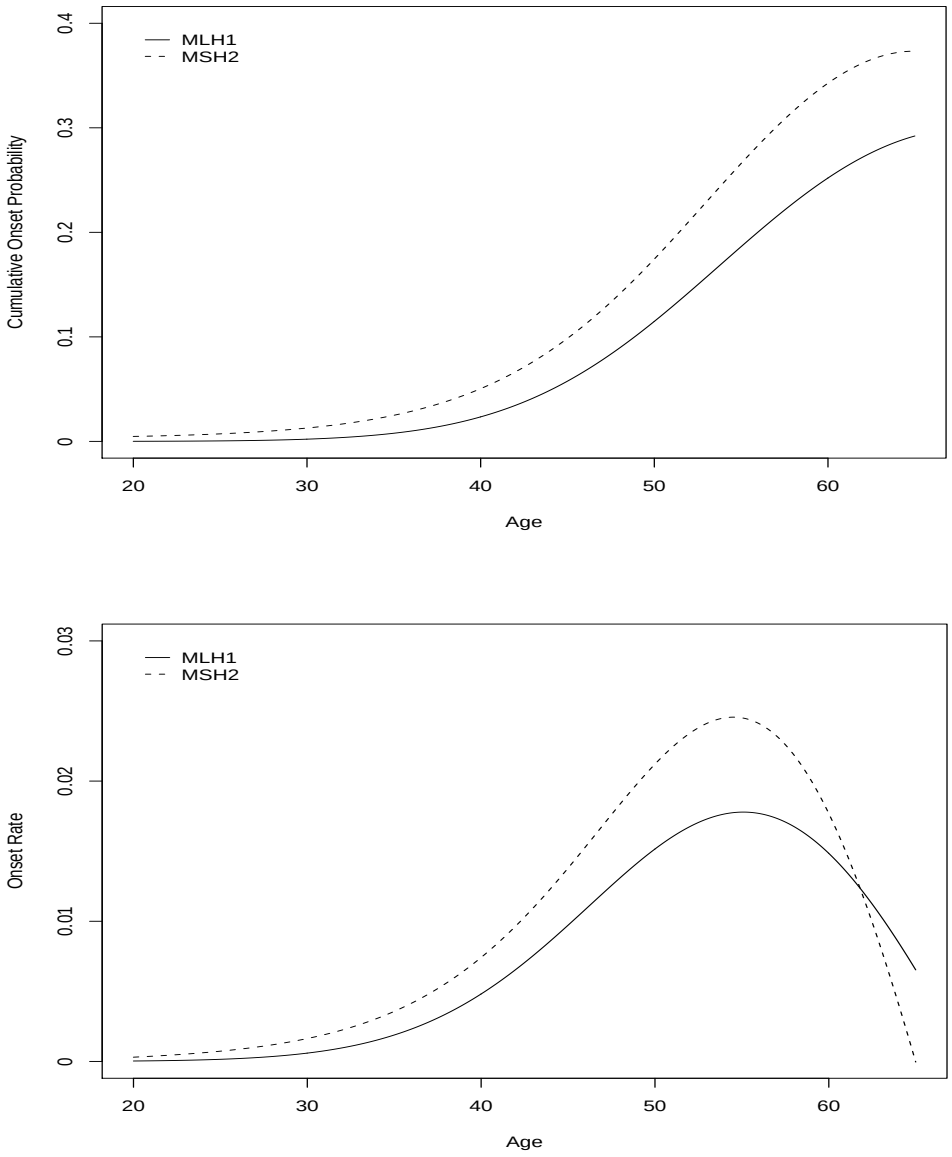

Figure 3.11: Fitted cumulative onset probability functions (top) and onset rates (bottom) for female MLH1 and MSH2 mutation carriers, in respect of EC. Source: Lu *et al.* (2007).

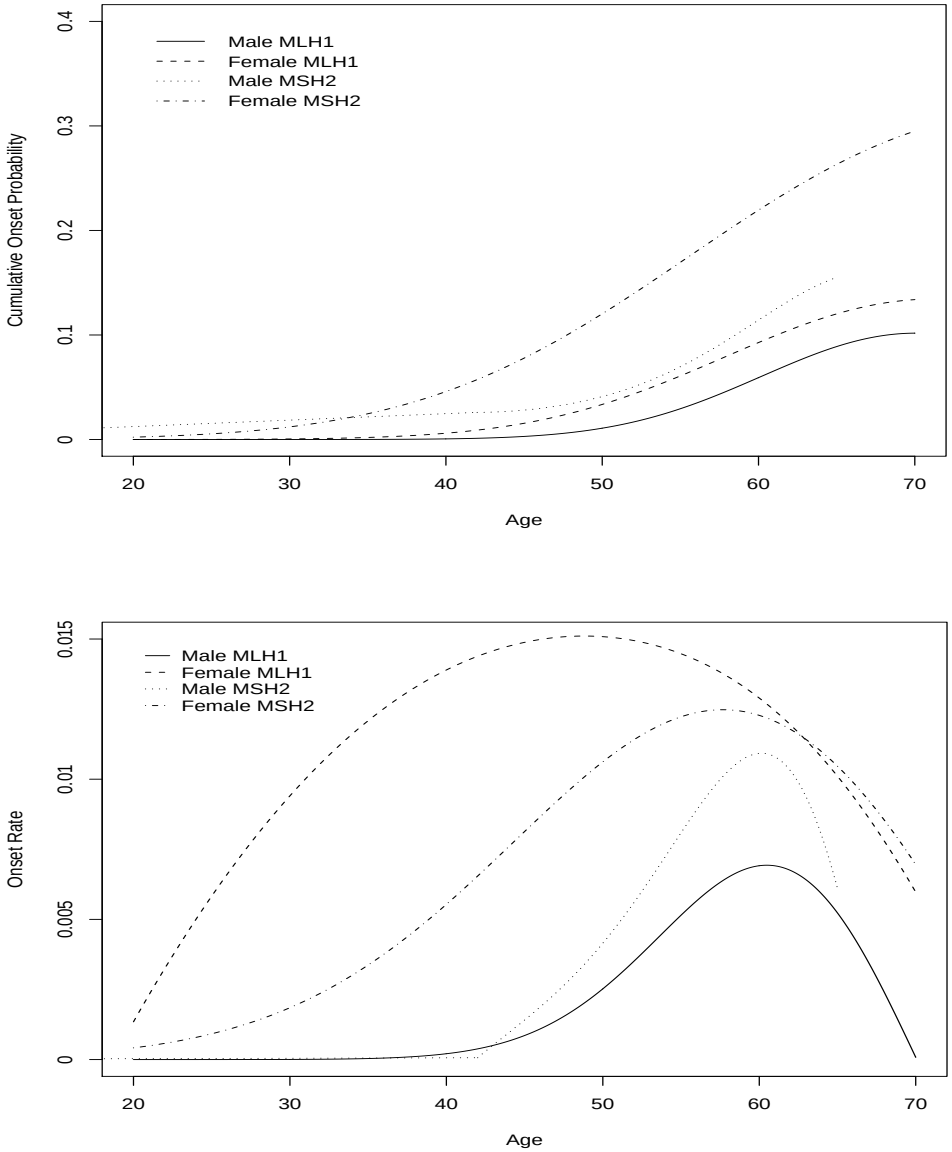

Figure 3.12: Fitted cumulative onset probability functions (top) and onset rates (bottom) for both male and female MLH1 and MSH2 mutation carriers, in respect of OECC. Source: Lu *et al.* (2007).

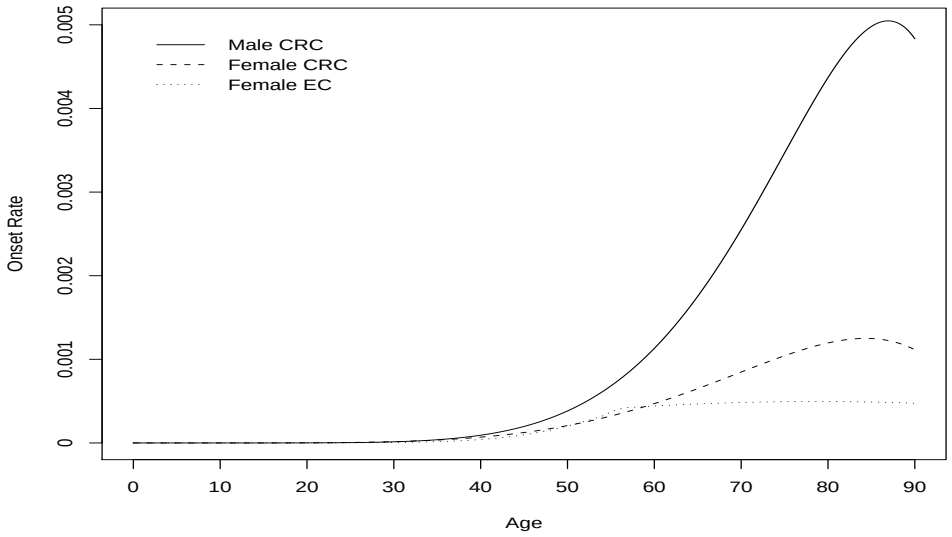

Figure 3.13: Fitted population onset rates of CRC for males and females and fitted population onset rates of EC for females. Source: Lu *et al.* (2007).

Post-Onset Mortality

Since mutations in MLH1 and MSH2 lead to the onset of cancers of several different sites (CRC, EC (female only) and OECC), the mortality after people develop any one of these three cancers needs to be estimated individually. In addition, the mortality experiences of HNPCC patients and sporadic CRC patients are different, so we need to estimate the post-onset mortality for both HNPCC and sporadic CRC patients.

(a) Post-onset mortality associated with CRC

We surveyed a number of papers on this topic. For brevity, please see the Appendix C.4 for more details. We chose Watson *et al.* (1998) as our main data source because this paper is one of the most comprehensive studies on post-CRC mortality, and we obtained the orginal dataset from Dr. Patrice Watson, the principal author of Watson *et al.* (1998). In Watson *et al.* (1998), this data was used to validate the hypothesis that patients with HNPCC have a better prognosis than sporadic CRC patients. Every patient in this data was staged according to the TNM system, and survival probabilities were estimated using Kaplan-Meier methods according to stage. In Chapters 3 and 4, the purpose of our model is to quantify the cost of adverse selection, and the modelling of HNPCC is only one part of the model, therefore we do not explicitly model staging.

In Watson *et al.* (1998), HNPCC cases, as the study group, were selected from 98 HNPCC families in the registries at Roswell Park Cancer Institute and Creighton University. The Amsterdam criteria were used to ascertain the HNPCC patients. As the control group, sporadic CRC cases were selected from the tumor registry (TR) of a single hospital affiliated with Creighton University. The diagnosis of CRC is taken as the start of survival analysis and the observation ends either by censoring because of death, or 10 years after the analysis. Kaplan-Meier methods were used

for the survival analysis.

Figure 3.15 shows the Kaplan-Meier estimate of the survival probability after contracting CRC, with our fitted curve (top) and the corresponding intensity $\mu_z^{CRC,Mortality,MC}$ (bottom) for HNPCC patients. Figure 3.14 shows the Kaplan-Meier estimate of the survival probability after contracting CRC, with our fitted curve (top) and the corresponding intensity $\mu_z^{CRC,Mortality,NC}$(bottom) for sporadic CRC patients. The superscripts MC and NC stand for mutation carriers and non-mutation carriers. Equation 3.1 is the function of duration since onset of CRC fitted to the post-onset survival probability associated with CRC for HNPCC patients and Equation 3.2 shows that for sporadic CRC patients. As in APKD and others diseases of interest, the post-CRC mortality rates should not be lower than that of $\mu_x^{Standard}$, e.g. English Life Tables 15. For mutation carriers:

$$\begin{aligned} S_z^{CRC,Mortality,MC} &= 1 - 0.05973z + 0.003277z^2 \\ &\quad -9.982 \times 10^{-5} z^3 + 1.035 \times 10^{-6} z^4, \end{aligned} \tag{3.1}$$

and:

$$S_z^{CRC,Mortality,NC} = \begin{cases} 1 - 0.1893z + 0.02413z^2 - 0.001342z^3 & z \leq 15 \\ 21.32 \cdot \frac{0.03250^{1.099}}{\Gamma(1.099)} \cdot z^{0.09905} \cdot \exp(-0.03250z) & z \geq 20 \end{cases} \tag{3.2}$$

Between $15 < z < 20$, we use a sine blending function.

We have the following comments:

(a) For mutation carriers, since data after duration 30 is very sparse, we should expect a substantial increase of the mortality rate after duration 30. We assume that the mortality rate levels off after duration 30.

(b) For non-mutation carriers, we found that it is hard to fit the Kaplan-Meier estimate of the survival probability after contracting CRC with one single function, because of the irregularity of the Kaplan-Meier estimate. Therefore, we use two functions to fit the Kaplan-Meier estimate. By trial and error, we fit the the part I (duration between 0 and 15 years) with a polynomial function and the part II (duration beyond 20 years) with the Gamma function. For duration between 15 and 20 years, we use a sine blending function.

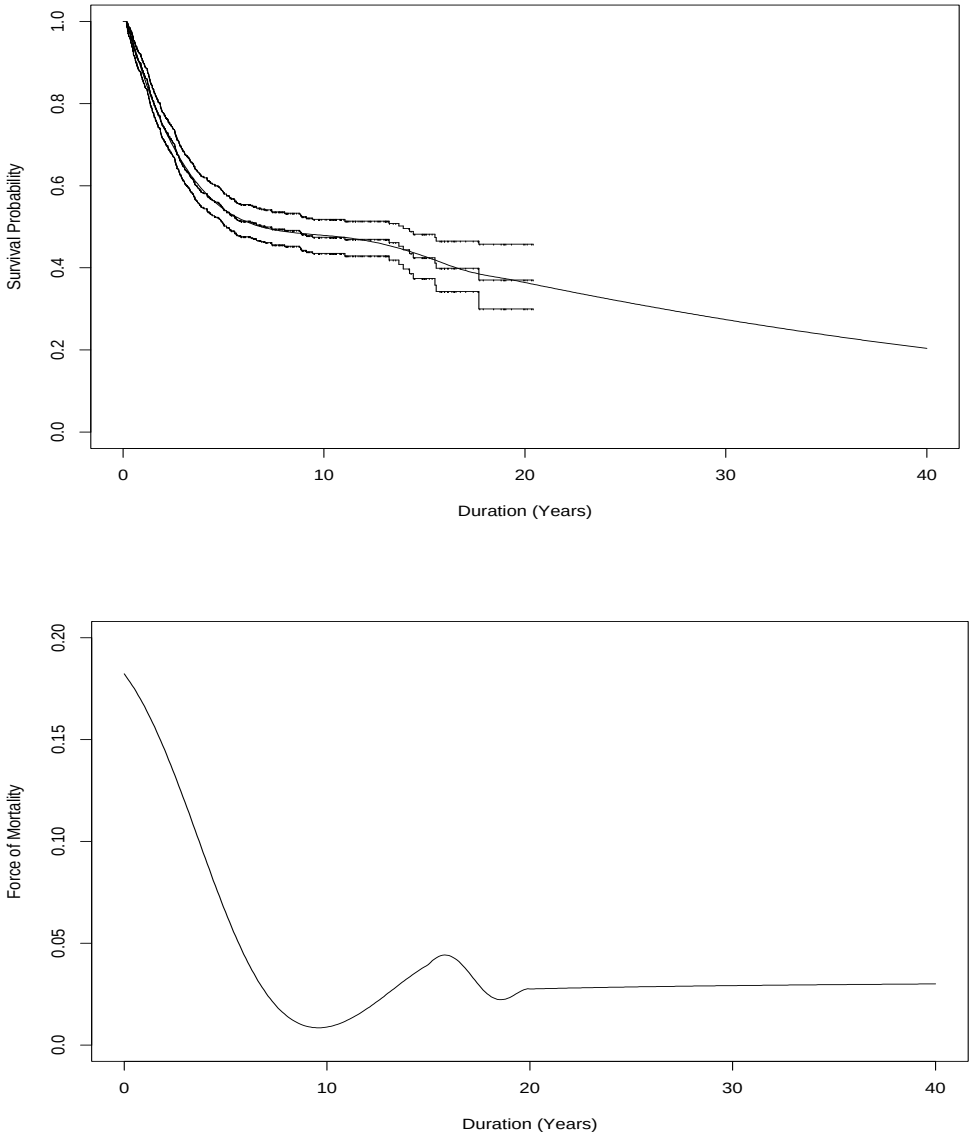

Figure 3.14: Kaplan-Meier estimates of survival probability after contracting CRC with fitted curve (top) and mortality intensity derived (bottom) for non-mutation carriers. Data source: Watson *et al.* (1998).

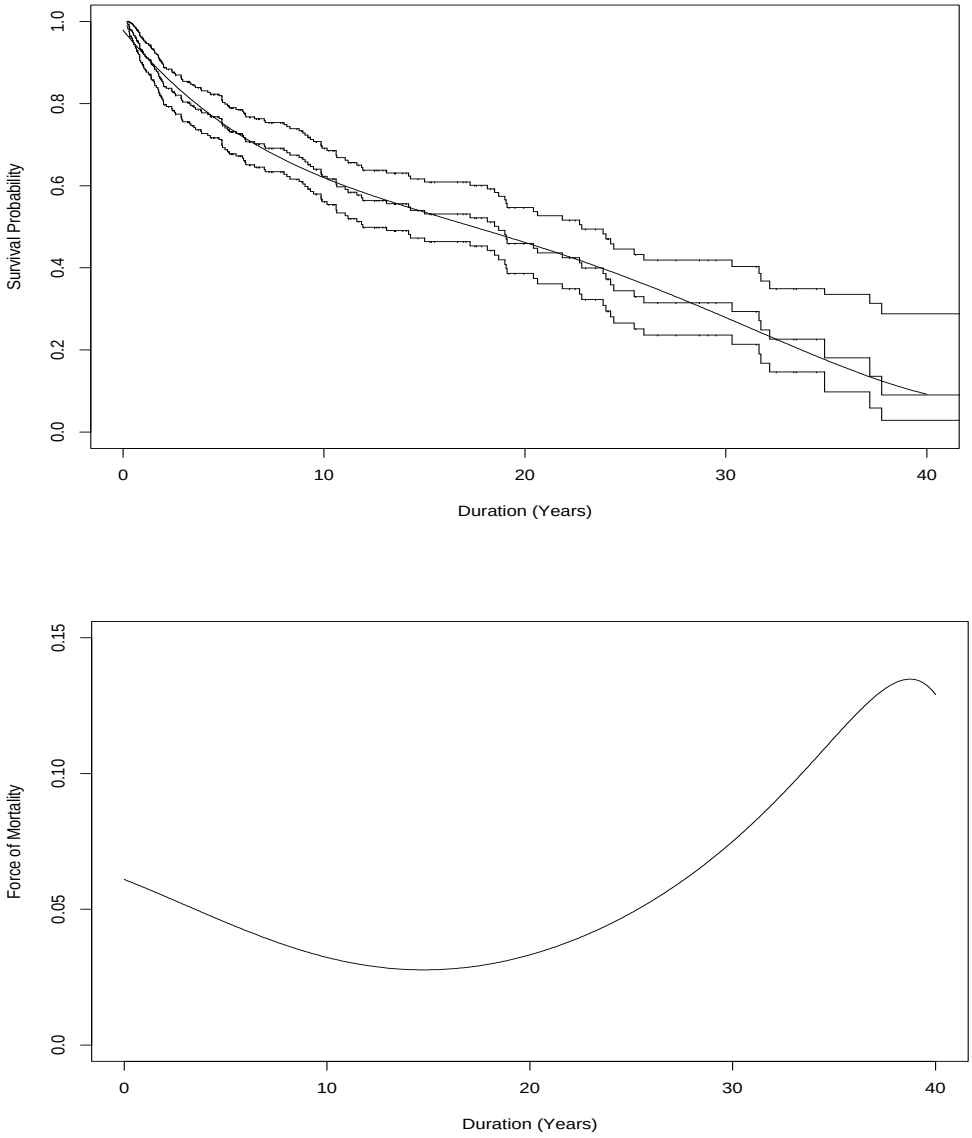

Figure 3.15: Kaplan-Meier estimates of survival probability after contracting CRC with fitted curve (top) and mortality intensity derived (bottom) for mutation carriers. Data source: Watson *et al.* (1998).

(b) Post-onset mortality associated with EC.

We reviewed several papers on this topic. Unfortunately, none of these papers carried out any survival analysis, which could be used in our work, except Boks *et al.* (2002). This might be because of the rarity of data compared with CRC. Boks *et al.* (2002) is a case-control study. The study group consisted of 50 patients with HNPCC-associated EC from the registry of the Netherlands Foundation for Hereditary Tumors. The control group consisted of 100 patients with sporadic EC registered in the Eindhoven Cancer Registry in the Netherlands. The observation time of the survival analysis was from the date of diagnosis until death or the end of the study in December 2000. The probability of survival since onset of EC was estimated using Kaplan-Meier methods. In the onset intensity analysis of Vasen *et al.* (2001), the observation time was from birth until date of diagnosis of cancer, death, or to the end of the study on July 1st, 2000. So the starting time of the survival analysis after onset of EC matches the ending time of the onset analysis with no time gap.

The conclusion is that the cumulative survival probability of patients with EC from HNPCC families is not significantly different from that of patients with sporadic EC. The disadvantage of this study is that the length of follow-up was short, 5 years for the control group and 10 years for study group, and the confidence intervals were not shown in the diagram. We used truncated Gamma functions to fit the Kaplan-Meier estimates and extrapolate the results to duration 40 years, because we assume in this thesis that the life insurance market operates between age 20 and 60. The extreme length of extrapolation is due to rarity of data. Figure 3.17 shows the Kaplan-Meier estimate of the survival probability after contracting EC, with our fitted curve (top) and the corresponding intensity $\mu_z^{EC,Mortality,MC}$(bottom) for HNPCC patients. Figure 3.16 shows the Kaplan-Meier estimate of the survival

probability after contracting EC, with our fitted curve (top) and the corresponding intensity $\mu_z^{EC,Mortality,NC}$(bottom) for sporadic EC patients. Equation (3.3) is the mathematical function fitted to the post-onset survival probability associated with EC for HNPCC patients and Equation 3.4 is that for sporadic EC patients. As in CRC, $\mu_z^{EC,Mortality,MC}$ and $\mu_z^{EC,Mortality,NC}$ should be assumed not to be lower than the population mortality in the ELT15.

$$1 - S_z^{EC,Mortality,MC} = \frac{0.004918^{0.6083}}{\Gamma(0.6083)} \int_0^z t^{-1.608} \exp(-0.004918t)dt \tag{3.3}$$

$$1 - S_z^{EC,Mortality,NC} = \frac{0.01036^{0.6492}}{\Gamma(0.6492)} \int_0^z t^{-1.649} \exp(-0.01036t)dt \tag{3.4}$$

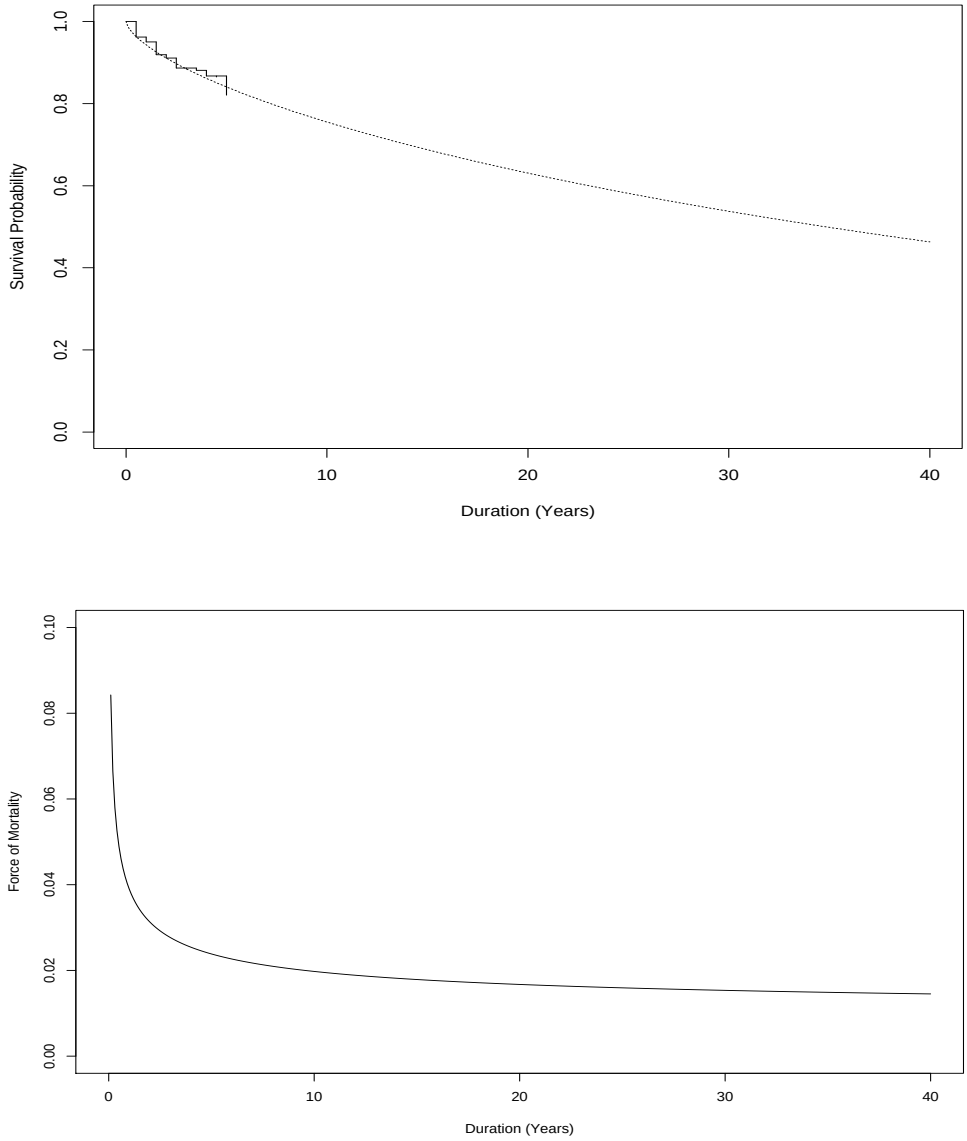

Figure 3.16: Kaplan-Meier estimate of the survival probability after contracting EC, with our fitted curve (top) and the corresponding mortality intensity (bottom) for sporadic EC patients. Data source: Boks *et al.* (2002)

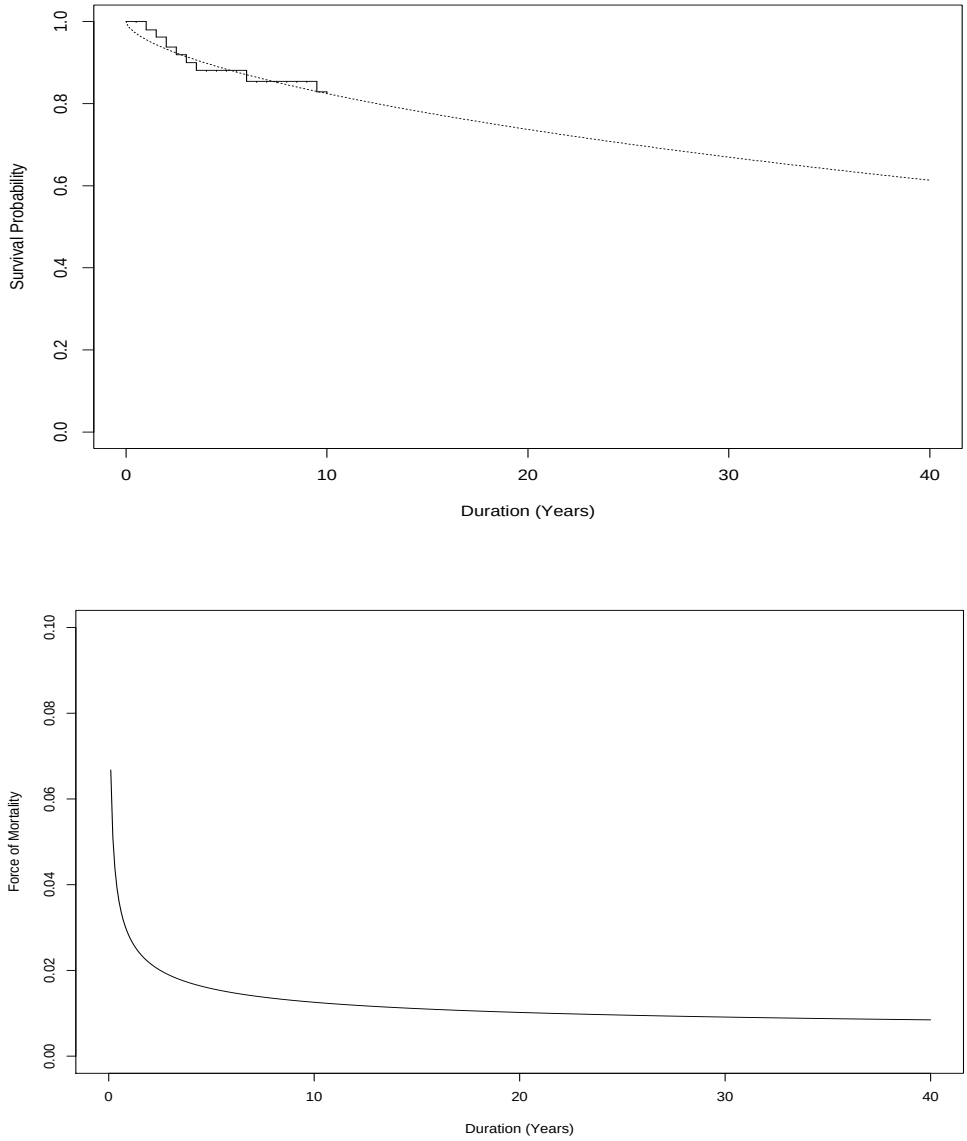

Figure 3.17: Kaplan-Meier estimate of the survival probability after contracting EC, with our fitted curve (top) and the corresponding mortality intensity (bottom) for HNPCC patients. Data source: Boks *et al.* (2002)

(c) Post-onset mortality associated with OECC.

OECC includes cancer of the stomach, urinary tract, small bowel, ovarian (female only) and brain. For the post-onset force of mortality associated with OECC, ideally we need to study the epidemiological literatures for each type of cancer, as in the case of CRC and EC. However, due to the rarity of HNPCC-associated OECC, no survival analysis study has been carried out in respect of some HNPCC-associated cancers (e.g. urinary tract cancer and brain cancer), and in some other cases (e.g. stomach cancer, small bowel cancer and ovarian cancer), the reliability of results is in doubt due to the small number of samples. We include a list of papers in Appendix C.4. Let $\mu_z^{OECC,Mortality,NC}$ and $\mu_z^{OECC,Mortality,MC}$ be the post-onset force of mortality associated with OECC for sporadic patients and HNPCC patients, respectively. In this thesis, for simplicity, we assume that $\mu_z^{OECC,Mortality,NC} = \mu_z^{CRC,Mortality,NC}$ and $\mu_z^{OECC,Mortality,MC} = \mu_z^{CRC,Mortality,MC}$. Considering the relatively low onset rates of OECC, we shall believe this assumption should have a very limited effect on results. As usual, the post-onset mortality associated with OECC should not be assumed to be lower than that of ELT15.

The Mortality Excluding Death Caused by CRC, EC and OECC

In the cases of single-gene disorders, we normally assume that they have negligible effect on overall mortality. But in the case of multifactorial disorders, e.g. HNPCC and BC & OC, this is not the case. Therefore, we need to calculate the rate of mortality excluding deaths caused by CRC, EC and OECC.

We calculated the ratio (r_x^{HNPCC}) of the numbers of deaths caused by CRC, EC and OECC to the total population numbers of deaths during 1990–1992 for each age x using the O.N.S. data. The force of mortality excluding deaths caused by CRC, EC and OECC is calculated as $\mu_x^{Standard} \times (1 - r_x^{HNPCC})$, where the population mortality $\mu_x^{Standard}$ is that

of ELT15. For males, the fitted function for r_x^{HNPCC} is

$$\begin{aligned} r_x^{HNPCC} &= -0.006208 + 0.004203x - 0.0004570x^2 \\ &\quad +2.152 \times 10^{-5}x^3 - 4.200 \times 10^{-7}x^4 + 3.613 \times 10^{-9}x^5, \end{aligned} \tag{3.5}$$

and for females, the fitted function for r_x^{HNPCC} is

$$\begin{aligned} r_x^{HNPCC} &= -2.308 + 0.360524x - 0.02261x^2 + 7.450 \times 10^{-4}x^3 - 1.380 \times 10^{-5}x^4 \\ &\quad +1.443 \times 10^{-7}x^5 - 7.960 \times 10^{-10}x^6 + 1.801 \times 10^{-12}x^7. \end{aligned} \tag{3.6}$$

Figure 3.18 shows the ratio r_x^{HNPCC} of the numbers of deaths caused by CRC, EC and OECC to the numbers of deaths in the population during 1992–1992, and fitted curves for males (top) and females (bottom). Figure 3.19 shows the adjusted mortality intensity $\mu_x^{Standard} \times (1 - r_x^{HNPCC})$ for males (top) and females (bottom).

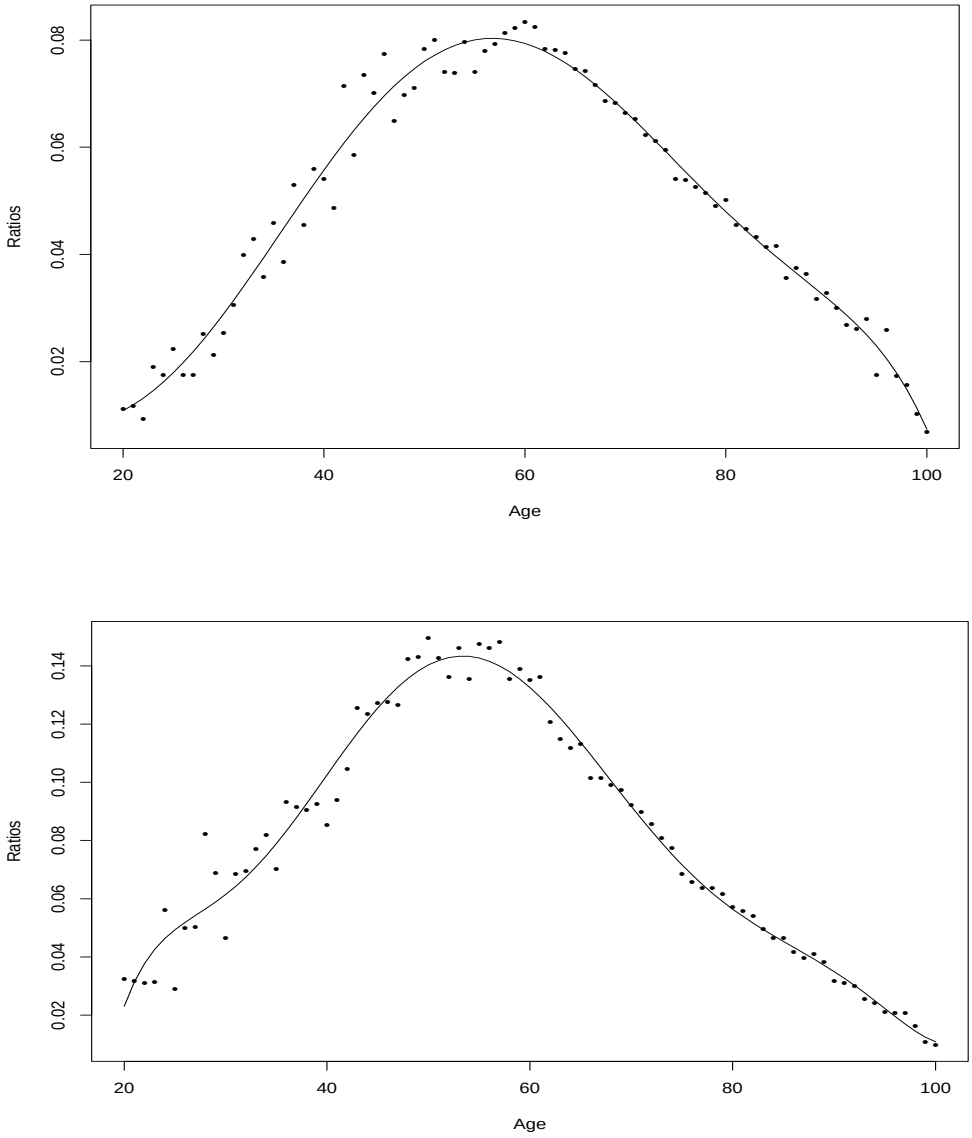

Figure 3.18: The ratio of the numbers of deaths caused by CRC, EC and OECC to the number of deaths in the population during 1992–1992, and fitted curves for males (top) and females (bottom)

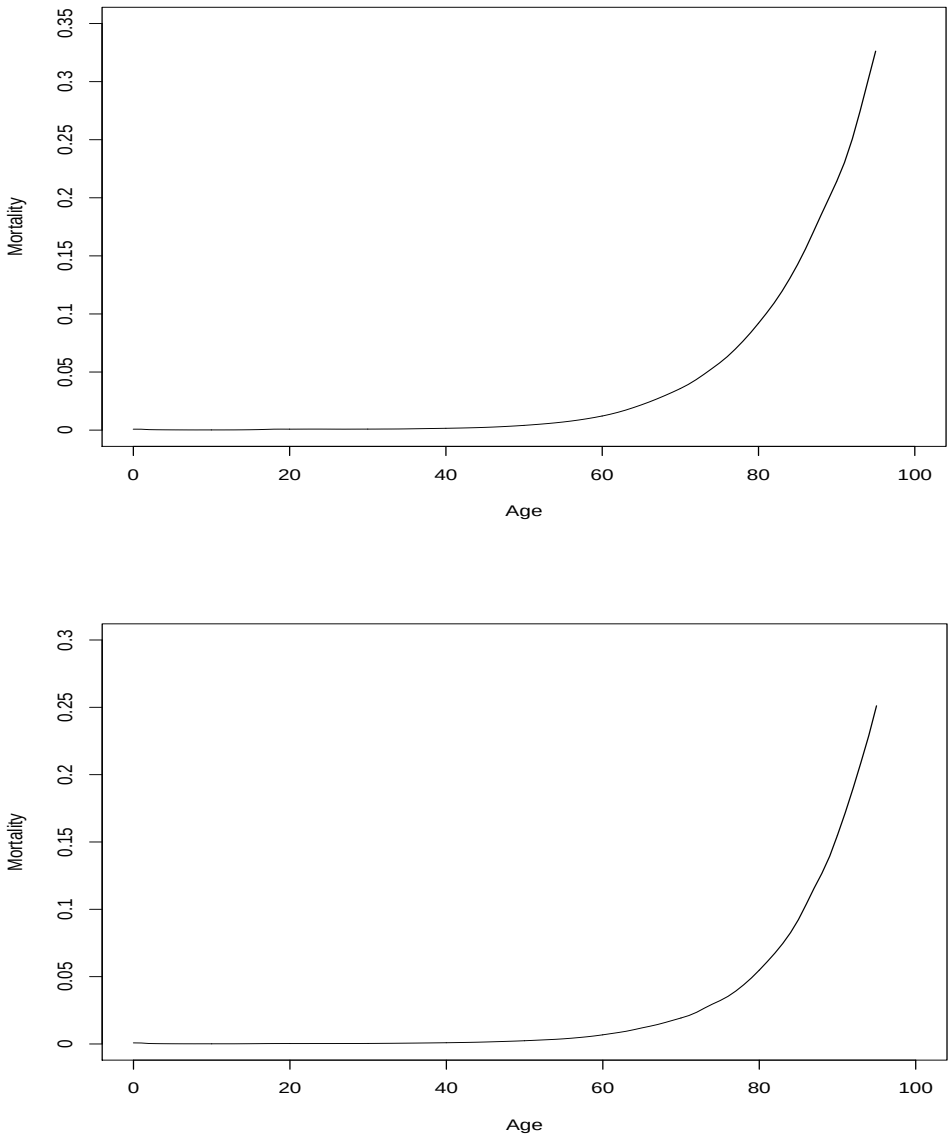

Figure 3.19: The adjusted mortality intensity excluding death caused by CRC EC and OECC for males (top) and females (bottom).

Prevalence Rate and the Distribution of Mutations

Lu *et al.* (2007) mentioned that Aaltonen *et al.* (1998) found that 2% of all CRC patients in Finland have HNPCC mutations, and that the Finnish Cancer Registry also found the risk of CRC by age 85 is 5% in the general population. Therefore Lu *et al.* (2007) assumed that the population frequency of HNPCC is about 1/1000. MLH1 and MSH2 mutations account for 90% of all HNPCC cases (Lynch & de la Chapelle, 2003). They assumed that the two mutations have roughly equal frequencies and the other minor MMR mutations have similar prognoses to those of MLH1 and MSH2 mutations. The penetrance of CRC in respect of MLH1 and MSH2 mutation carriers is estimated at about 65% and 70% by age 75 years in Lu *et al.* (2007), therefore Lu *et al.* (2007) took the population frequencies of MLH1 and MSH2 mutations to be 0.0005 / 0.65 = 0.000769 and 0.0005 / 0.7 = 0.000714 respectively.

The Onset Rate of a Family History

In the case of diseases with purely genetic causes, e.g. APKD, EOAD, HD and MD, because the lifetime penetrance is close to 100%, persons at risks of single-disorders are assumed to have a family history at outset, which is taken to mean an affected parent. However, in the case of HNPCC and BC & OC, this is not the case, i.e. a family history might arise by chance in any family.

Under the family history underwriting, we need to estimate the conditional probability that a person aged x has a given genotype g say, given their family history at that age:

$$\mathbf{P}\left[\text{genotype } g_i \mid \text{healthy at age } x \text{ and with family history}\right],$$

In the case of CRC which have both hereditary and sporadic causes, this probability is difficult to estimate, because it can depend on other family members, i.e. on family structure, and the definition of family history. A family history relevant to CRC has been

defined in underwriting practice. Lu *et al.* (2007) considered three scenarios that might represent underwriting criteria:

(a) Scenario 1: at least one first degree relative (FDR) develops CRC before age 60;

(b) Scenario 2: at least two FDRs develop CRC before age 60;

(c) Scenario 3: at least one FDR develops CRC before age 60 and at least one FDR develops EC before age 60.

This suggests that, given an explicit definition of a family history, we can model its appearance as an event, or a transition between two states. We use one example to demonstrate this idea. Suppose we use scenario 2 as the relevant threshold for underwriting. If at age $x-1$ a woman had one affected parent but by age x she had one affected parent and one affected sister, then this woman crossed that threshold between age $x-1$ and age x. We can see from this example that the onset rate of a family history partly depends on the structure of a family, which needs to be modeled separately. A representative distribution of the number of siblings can be found in Macdonald, Waters & Wekwete (2003a) (see Table 3.1).

Table 3.1: The distribution of the number of siblings. Source: Macdonald, Waters & Wekete (2003a).

No. of siblings k	0	1	2	3	4	5	≥ 6
Probability $\mathbf{P}[k]$	0.2311	0.5006	0.1973	0.0541	0.0124	0.0034	0.0011

Based on this information, we can then calculate the onset rate of a family history numerically. In this thesis, for simplicity, we choose to use only scenario 2 as the definition

of a family history relevant to CRC. Taking scenario 2 as an example, we briefly introduce the methodology of production of the onset rate of a family history as follows:

1. Sub-populations

   The entire population is partitioned into five sub-populations according to a person's own genotype and the presence of a mutation in their family as follows:

   (a) Sub-population 1: not carriers of any MMR gene mutation and have no MMR gene mutation carriers within their family;

   (b) Sub-population 2: MLH1 mutation carriers and have MLH1 mutation carriers within their family;

   (c) Sub-population 3: not MLH1 mutation carriers but in an MLH1 mutation carrier family;

   (d) Sub-population 4: MSH2 mutation carriers and have MSH2 mutation carriers within their family;

   (e) Sub-population 5: not MSH2 mutation carriers but in an MSH2 mutation carrier family;

   Every one is at risk of CRC (either hereditary or sporadic) no matter which sub-population (s)he is in. The proportion of the population in sub-population 2 and 4 are the frequencies of the MMR mutations among the general population, which were taken to be 0.000769 and 0.000714, respectively. A new-born baby of a mutation carrier has probability 1/2, at birth, of being a carrier. Hence the proportion in Sub-populations 3 and 5 are the same as those in sub-populations 2 and 4, respectively.

2. The Model

   We use a Markov model in Figure (3.20) to describe the life history of a person at risk of CRC, EC and OECC and incorporate the feature of onset of family history.

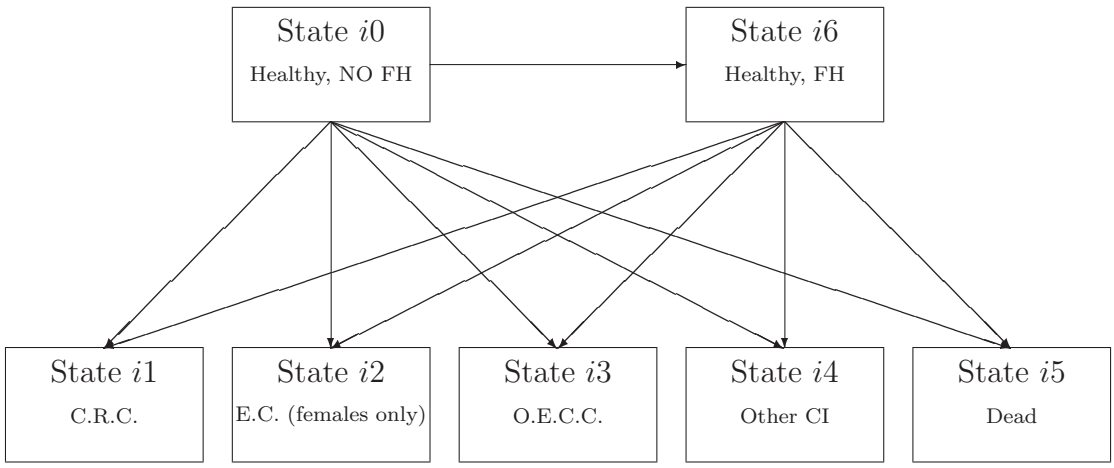

Figure 3.20: A life history model of a person at risk of CRC, EC and OECC, incorporating the development of a family history, in sub-population i.

The onset rate of a family history is intensity μ_x^{i06} in Figure (3.20). We also assume that for each person in sub-population i, since their mutation status is known, the onset rates of CRC, EC and OECC are fully defined in Section 3.4.2, which are unaffected by the presence of absence of a family history (Lu *et al.*, 2007). Let OP_x^{i6} be the probability of being State $i06$ at age x for a person in sub-population i. We have:

$$\mathbf{P}\left[\text{genotype } g_i \mid \text{healthy at age } x \text{ and with family history}\right] = \frac{OP_x^{i6}}{\sum_{j=1}^{5} OP_x^{j6}}. \quad (3.7)$$

3. Assumptions

In the calculation of the onset rate of a family history, we made the following assumptions (Lu *et al.*, 2007):

(a) The parents are 30 years older than the applicant and any siblings are all the same age as the applicant.

(b) The distribution of the number of children is given in Table 3.1.

(c) Siblings could be male or female with equal probability.

(d) The siblings in a mutation carrier family (i.e. sub-populations 2 to 5) have equal probability of being a carrier or a non-carrier. The probability that such people develop cancer is an average over that for carriers and non-carriers.

(e) Only one parent is a mutation carrier, but the other one not. We neglect the possibility that both parents are mutation carriers.

4. Calculation

We adopted the notation system from Lu *et al.* (2007). Let ${}^iP_x^{0P}$, ${}^iP_x^{1P}$ and ${}^iP_x^{2P}$ denote the probabilities that neither of their parents, one of their parent or both of their parents, respectively, suffer onset of CRC by age $x + 30$. Let ${}^iP_x^{0S}$, ${}^iP_x^{1S}$ and ${}^iP_x^{2S}$ denote the probabilities that they have none, one or two siblings, respectively, suffering onset of CRC by age x. Given a person in sub-population i, we can estimate the probability that they have a family history by age x, denoted as ${}^iF_x^{fh}$, under scenario 2, which is:

$$ {}^iF_x^{fh} = {}^iP_x^{0P} \times (1 - {}^iP_x^{0S} - {}^iP_x^{1S}) + {}^iP_x^{1P} \times (1 - {}^iP_x^{0S}) + {}^iP_x^{2P}. \tag{3.8} $$

By numerical differentiation, we can then obtain the intensity μ_x^{i06} in Figure (3.20), the rate of onset of a family history. For more calculation details, please see Lu *et al.* (2007).

5. Some features

There is a downward jump at age 30, because the parents then reach age 60 and can no longer contribute to a family history. The rate of 'onset' for MLH1 mutation carriers jumps at about 20, because the rate of onset of CRC for MLH1 mutation carriers is assumed to be zero before age 22 (Lu, 2006).

Figure 3.21 shows the onset rates of a family history for non-mutation, MLH1 mutation and MSH2 mutation families under scenario 2.

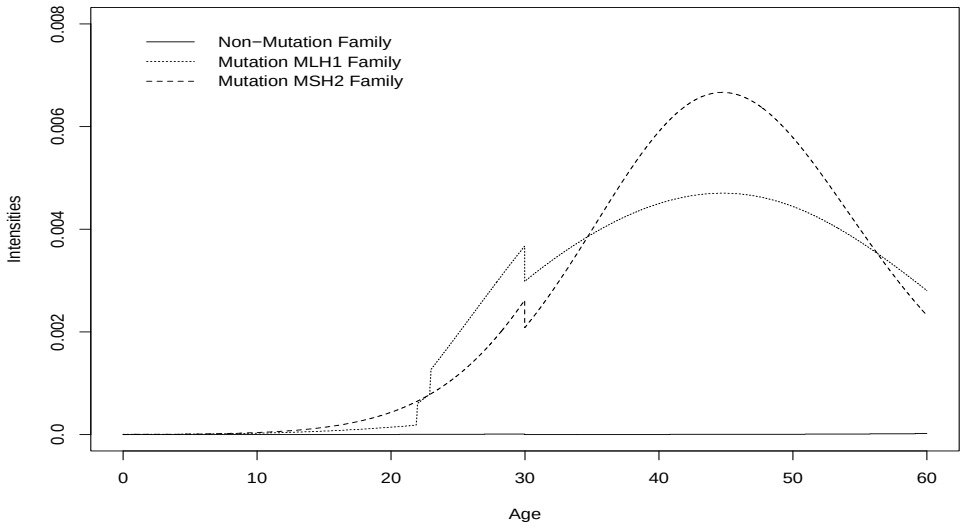

Figure 3.21: The onset rates of a family history for non-mutation, mutation MLH1 and mutation MSH2 families.

3.5 Breast Cancer and Ovarian Cancer

3.5.1 Introduction to Breast Cancer and Ovarian Cancer

Among women worldwide, breast cancer is the most common cancer. Breast cancer (BC) refers to a malignant tumor that has developed from cells in the breast. Several factors are recognized as contributing to an increased risk of developing BC, including age, region, breast density and family history. The possibility may increase four- to six-fold if the woman has two First Degree Relatives diagnosed with the disease (Claus *et al.*, 1996). So this leads epidemiologists to classify BC cases into two categories: sporadic forms and hereditary forms.

In the case of hereditary forms, mutations in two genes, BRCA1 and BRCA2, have been identified that predispose individuals to BC. Mutations in the BRCA1 and BRCA2 genes confer very high risk of breast cancer (BC) but only account for about 25% of the observed familial clustering of BC. Antoniou *et al.* (2001) proposed a model that included the BRCA1 and BRCA2 genes, and a polygenic component that acted multiplicatively on the rate of onset of BC.

Ovarian cancer (OC) is a disease in which malignant or cancerous cells are found in the ovary. OC is the second most common malignancy in women (after BC). There are several types of OC. The most common type is epithelial OC, which begins on the surface of the ovary, making up 90% of all OC cases. In a process called shedding, OC cells can break away from the ovary and spread to other tissues and organs. Shedding may make a woman feel bloated, or make her abdomen look swollen. However, these symptoms are easily missed because they can all be caused by a number of other diseases. This makes it difficult to diagnose OC by symptoms alone. At this moment, there is no reliable method of screening for OC. In clinical practice, both CA125 blood test and vaginal ultrasound are currently being applied as possible approaches of screening women for OC. Several

factors may increase developing the chance of developing OC, including age, genetics, family history, and fertility drugs.

Surgery is the preferred treatment and is frequently necessary for differential diagnosis. The type of surgery depends upon how widespread the cancer is when diagnosed, as well as the type and grade of cancer. The surgeon may remove one ovary (unilateral oophorectomy) or both ovaries (bilateral oophorectomy). For some very early tumors, only the involved ovary and fallopian tube will be removed (called a "unilateral salpingo-oophorectomy"), especially in young females who wish to preserve their fertility. In advanced disease as much tumor as possible is removed (debulking surgery). In cases where this type of surgery is successful, the prognosis is improved compared to patients where large tumor masses (more than 1 cm in diameter) are left behind. Chemotherapy is used after surgery to treat any residual disease, if appropriate.

3.5.2 Epidemiology of BC & OC

Onset Rates

For incidence rates of BC and OC associated with BRCA1 and BRCA2 mutations, a recent meta-analysis by Antoniou *et al.* (2001) pooled pedigree data from 22 studies including 8,139 index cases. Age-specific cumulative risks were estimated by Kaplan-Meier methods, and approximate piecewise constant rates were obtained. Gui *et al.* (2006) fitted truncated Gamma functions to the piecewise constant onset rates, using unweighted least squares. Please see Appendix C.5 for the fitted truncated Gamma functions. Figure 3.22 shows the fitted onset rates of BC (top) and OC (bottom) for BRCA1 and BRCA2 mutation carriers.

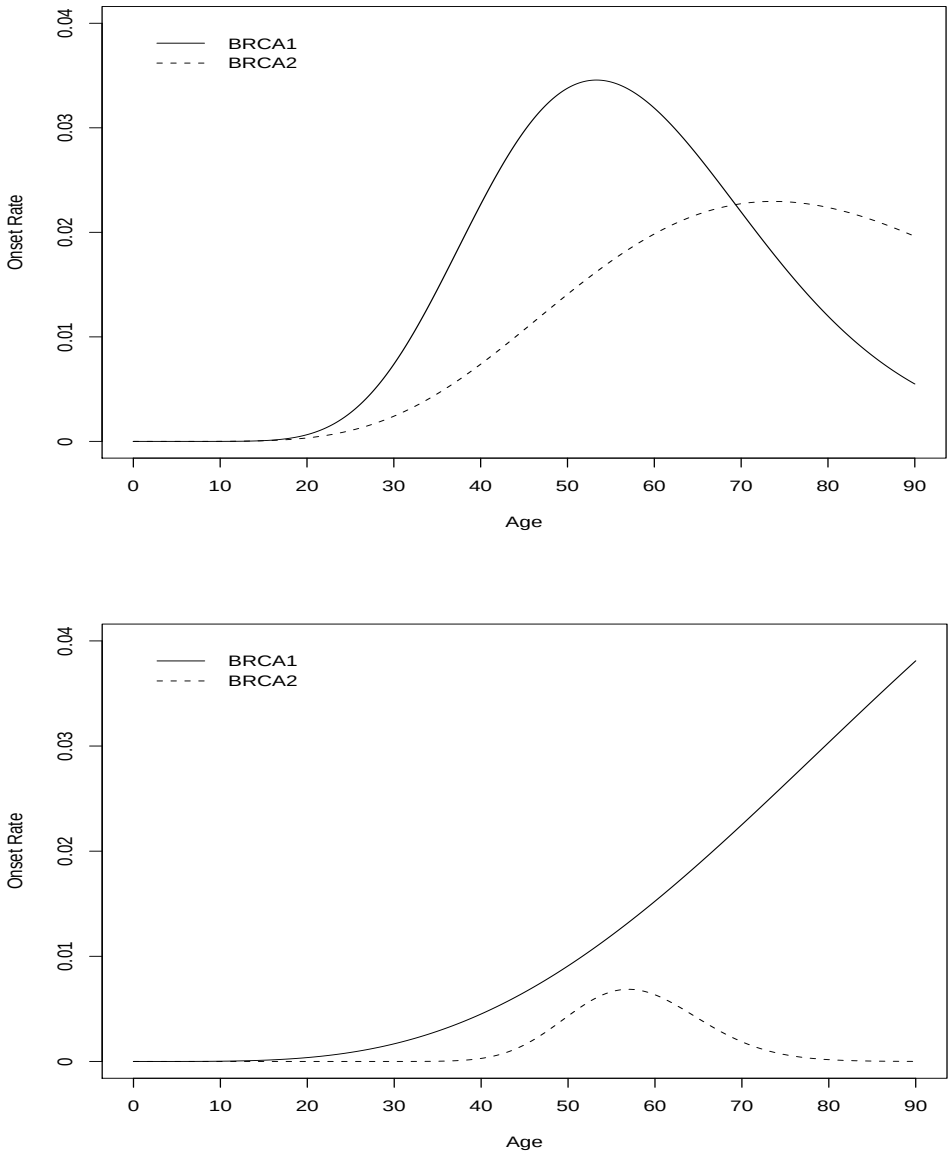

Figure 3.22: Fitted onset rates of BC (top) and fitted onset rates of OC (bottom) for BRCA1 and BRCA2 mutation carriers. Source: Gui *et al.* (2006).

As to the population onset rate of BC and OC, Wekwete (2002) estimated these using O.N.S. data from years 1984–1988. The crude rates $\dot{\mu}_x^{BC,POP}$ and $\dot{\mu}_x^{OC,POP}$ were calculated as:

$$\dot{\mu}_x^{BC,POP} = \frac{\theta_x^{BC,POP}}{{}^cE_x^{BC,POP}}, \tag{3.9}$$

and:

$$\dot{\mu}_x^{OC,POP} = \frac{\theta_x^{OC,POP}}{{}^cE_x^{OC,POP}}, \tag{3.10}$$

where $\theta_x^{BC,POP}$ and $\theta_x^{OC,POP}$ are the numbers of cases of BC and OC respectively, and ${}^cE_x^{BC,POP}$ and ${}^cE_x^{OC,POP}$ are the corresponding exposed to risk. The crude rates were then fitted using unweighted least squares. In order to differentiate the crude rates from the fitted rates, we denote the fitted rates $\mu_x^{BC,POP}$ and $\mu_x^{OC,POP}$ for BC and OC respectively. Please see Appendix C.5 for the mathematical expressions of $\mu_x^{BC,POP}$ and $\mu_x^{OC,POP}$. Figure 3.23 shows the fitted population onset rates for BC and OC. It can be observed that the BC incidence rates are unusual at age 50–64. This is because of the introduction of systematic BC screening in the UK in women bebetween the ages of 50 and 64.

Post-Onset Mortality Rate

Gui *et al.* (2006) studied the mortality rates after the onset of BC or OC. Please see Appendix C.5 for technical details. Figure 3.24 shows the mortality rates after the onset of BC (top) and OC (bottom) as a function of duration since onset.

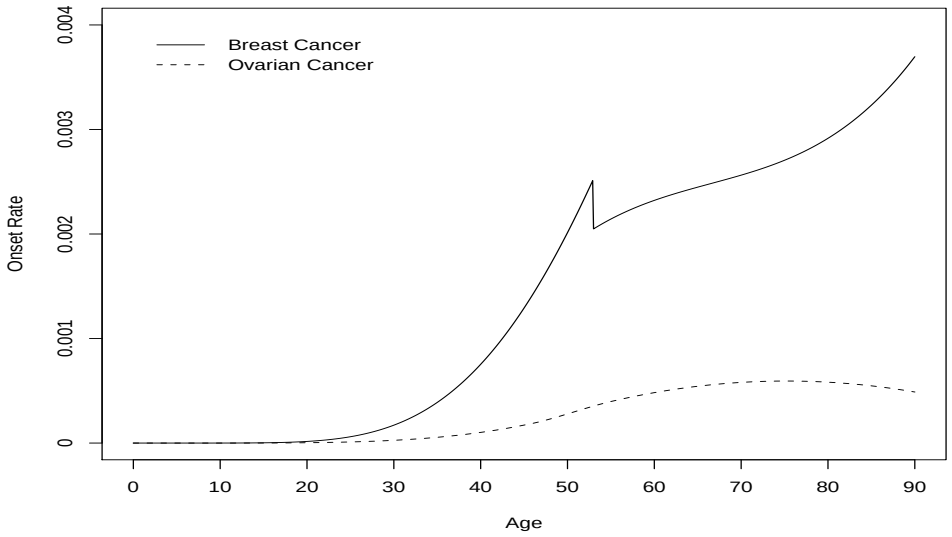

Figure 3.23: Fitted population onset rates of BC and OC. The unusual BC incidence rates at age 50–64 is because the introduction of systematic BC screening in the UK in women between the ages of 50 to 64. Source: Wekwete (2002).

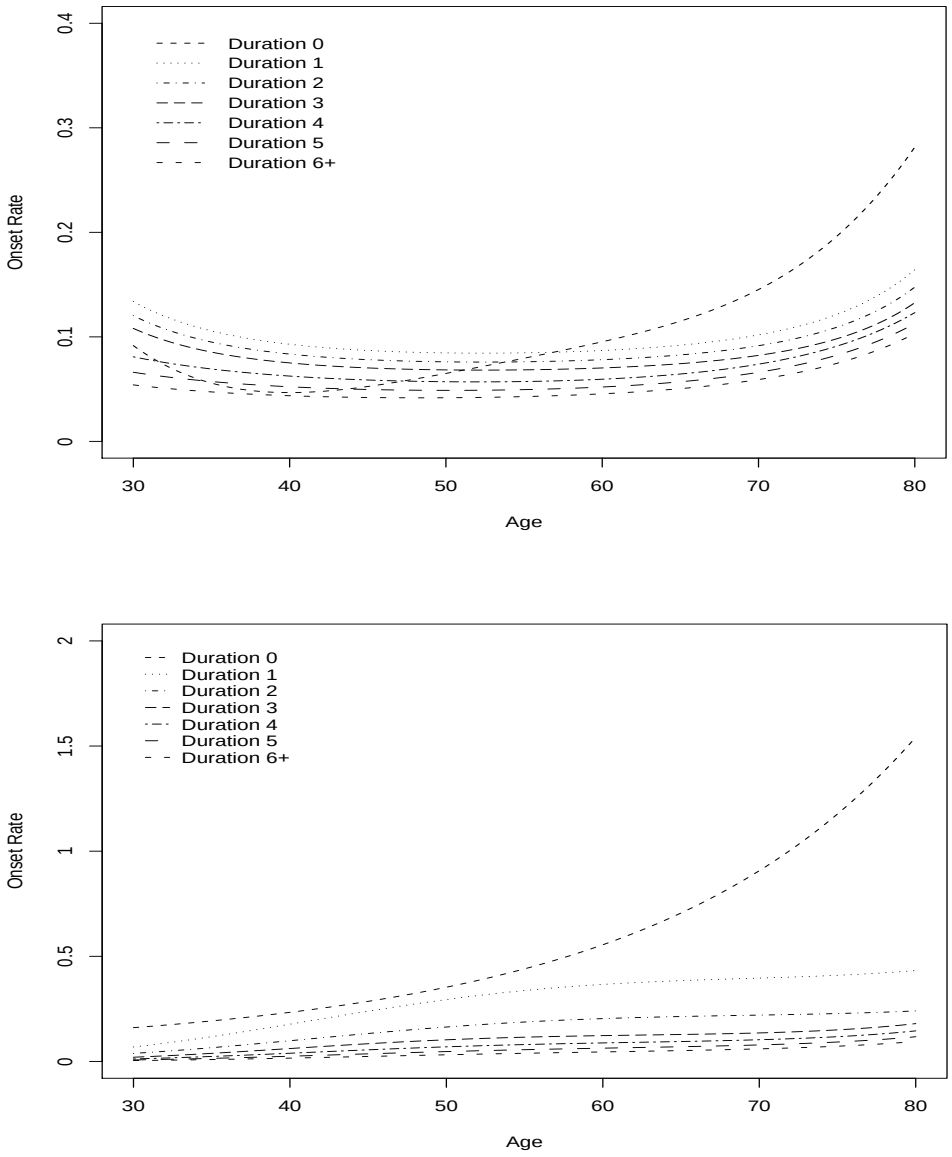

Figure 3.24: Fitted mortality rates after onset of BC (top) and OC (bottom) for all durations since onset. Source: Gui *et al.* (2006).

Mortality Excluding Death Caused by BC and OC

Wekwete (2002) calculated the ratios (r_x^{BCOC}) of the numbers of deaths caused by BC and OC during 1990–1992 to the total numbers of deaths during 1990–1992, for each age x using O.N.S. data. The force of mortality excluding deaths caused by BC and OC is calculated as $\mu_x^{Standard} \times (1 - r_x^{BCOC})$, where the population mortality $\mu_x^{Standard}$ is that of ELT15. Figure 3.25 (top) shows the curve fitted to the ratio r_x^{BCOC} of the number of deaths cause by BC and OC to the number of deaths in the population during 1992–1992 for females. Figure 3.25 (bottom) shows the corresponding adjusted mortality intensity $\mu_x^{Standard} \times (1 - r_x^{BCOC})$.

Prevalence Rate and the Distribution of Mutations

Antoniou *et al.* (2001) estimated allelic mutation frequencies of BRCA1 and BRCA2 mutations to be 0.0005829 and 0.000676 respectively.

The Onset Rate of a Family History

As in the case of HNPCC, a family history relevant to BC & OC can occur in any family. Therefore, we need to model the onset of a family history by making the appearance of a family history as a precisely defined event. Gui *et al.* (2006) chose the following underwriting threshold for a family history: two first-degree relatives suffer onset of BC or OC before age 50. Hence if by age $x - 1$ a woman has one sister suffering BC before age 50 and by age x she has two, then this woman crosses the threshold between age $x-1$ and x. The distribution of the number of the applicant's sisters is modeled in Macdonald, Waters & Wekwete (2003a). Hence the onset rate of a family history can be calculated straightforwardly. Please refer to Gui *et al.* (2006) for more detailed information. Figure 3.26 shows the onset rates of a family history for non-mutation, mutation BRCA1 and mutation BRCA2 families.

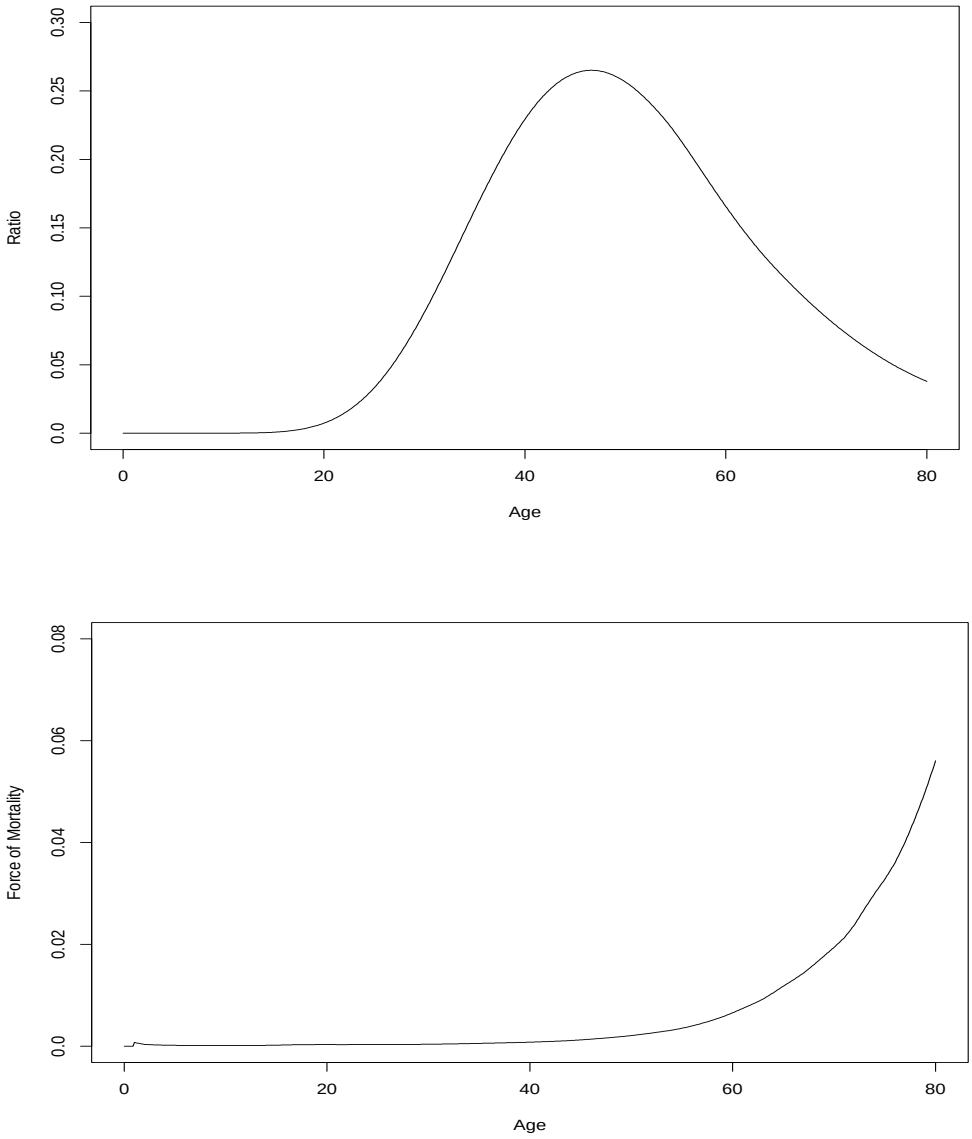

Figure 3.25: The curve fitted to the ratio of the number of deaths cause by BC and OC to the number of deaths in the UK population during 1992–1992 (top) and the corresponding mortality intensity adjusted to exclude the deaths caused by BC & OC (bottom). Source: Gui *et al.* (2006).

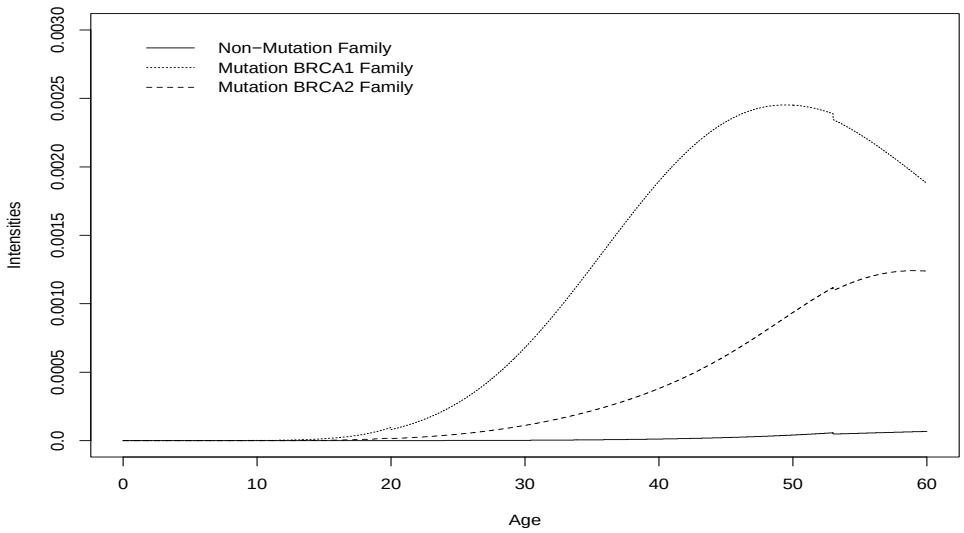

Figure 3.26: The onset rates of a family history for non-mutation, mutation BRCA1 and BRCA2 families. The family history is defined as two first-degree relatives suffering onset of BC or OC before age 50. Source: Gui *et al.* (2006).

Chapter 4

Overall Impact of Genetic Information - "Bottom-Up" Approach

In this chapter, we will look at the overall impact of genetic information, so as to come closer to completing the "bottom-up" approach. In individual studies, when we talk about the impact of genetic information, we look at two questions: the insurability of insurance applicants, and the cost of adverse selection. The first question has been fully answered in each individual study (see Section 1.2.5 for the list of individual studies). Here, we consider the second question, the cost of adverse selection.

4.1 Critical Illness Insurance Market Models

In Chapter 2, we studied the cost of adverse selection relating to MD in the CI insurance and life insurance markets. This individual study of MD exemplified the methodology to quantify the cost of adverse selection by modelling the whole insurance market using a

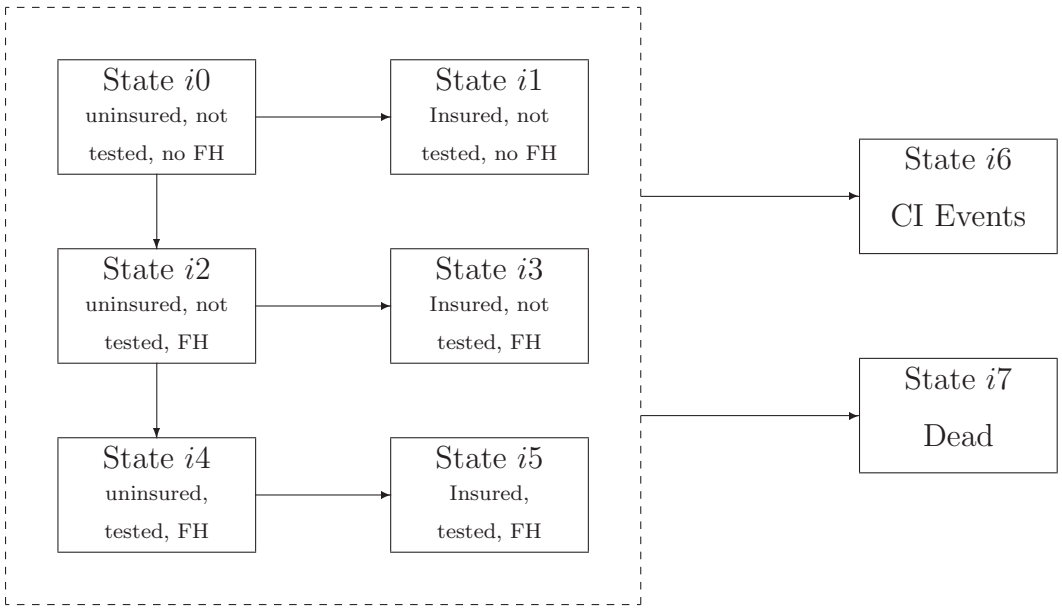

Figure 4.1: A Markov model of family history, genetic testing, insurance purchase and CI insurance events for a person in the i^{th} risk subpopulation (FH = family history present).

Markov or semi-Markov model. This market model includes several factors, e.g. insurance applicants' insurance-purchasing behavior, genetic testing, development of a family history, and more. Figure 4.1 shows a CI insurance market model, which we used to quantify the overall impact of genetic information.

In this thesis, we choose six genetic disorders, taken to have significant impact on the insurance industry (see Chapter 1 for details). Each of these genetic disorders may be caused by one or more mutations. Depending on whether a person carries a mutation

and/or a family history, we divide the whole population into the following sub-populations:

(a) sub-population 1: not mutation carriers, not having family history

(b) sub-population 2: carrying mutation APKD1, with family history

(c) sub-population 3: not carrying mutation APKD1, but with family history

(d) sub-population 4: carrying mutation APKD2, with family history

(e) sub-population 5: not carrying mutation APKD2, but with family history

(f) sub-population 6: carrying mutation PSEN-1, with family history

(g) sub-population 7: not carrying mutation PSEN-1, but with family history

(h) sub-population 8: carrying mutation CAG, with family history

(i) sub-population 9: not carrying mutation CAG, but with family history

(j) sub-population 10: carrying mutation CTG repeat 250+, with family history

(k) sub-population 11: not carrying mutation CTG repeat 250+, but with family history

(l) sub-population 12: carrying mutation CTG repeat 250–, with family history

(m) sub-population 13: not carrying mutation CTG repeat 250–, but with family history

(n) sub-population 14: carrying mutation MLH1, with family members being mutation carriers

(o) sub-population 15: not carrying mutation MLH1, with family members being mutation carriers

(p) sub-population 16: carrying mutation MSH2, with family members being mutation carriers

(q) sub-population 17: not carrying mutation MSH2, with family members being mtation carriers

(r) sub-population 18: carrying mutation BRCA1, with family members being mutation carriers

(s) sub-population 19: not carrying mutation BRCA1, with family member being mutation carriers

(t) sub-population 20: carrying mutation BRCA2, with family members being mutation carriers

(u) sub-population 21: not carrying mutation BRCA2, with family members being mutation carriers.

Some features of this model follow:

(a) The purpose of this model is to estimate the cost of adverse selection arising from a CI insurance market. We assume the CI insurance market operates between ages 20 and 60.

(b) Comparing with the market model in Figure 2.6, the new market model has two extra states: State 0, "uninsured, not tested, no FH", and State 1, "insured, not tested, no FH". The purpose is to consolidate the market model for diseases with no cause except mutations in single genes and single gene subsets of common disorders for the ease of completing the "bottom-up" program. Sub-populations 2–13 refer to the diseases with no cause except mutation in a single gene. Because of high penetrance, persons in these groups are assumed to have a family history, consisting of an affected parent. So the starting state for these populations is state 2, "No Insurance, with family history (FH)". Sub-populations 1 and 14–21 refer to single

gene subsets of common disorders. Persons in these groups start in State 0, "No Insurance, without family history (FH)".

(c) The proportion of the population in each mutation-carrying sub-population is given by the prevalence of the mutations, as stated in Chapter 3. The proportion in the corresponding sub-population of non-mutation carriers will be the same, since a mutation will be inherited from the affected parent with probability 1/2.

(d) Family history is an important underwriting factor in practice. In this model, for sub-populations 2–13, a family history is present from outset, so μ_x^{i02} is irrelevant, but for sub-populations 14–21, a family history may emerge later, so $\mu_x^{i02} > 0$, in general. A separate model was deployed to calculate this intensity numerically. Please refer to Sections 3.4.2 and 3.5.2 for more detailed information.

(e) A genetic test result is a source of asymmetric information between the insured and insurer. The intensity μ_x^{i24} represents the rate of uptake of testing. In this paper, the baseline rate of testing is 0.014 per annum between ages 20 and 40, meaning about that about 10% would be tested after 8 years, and the other two scenarios are 0.014 per annum between ages 20 and 60, and 0.035 per annum between ages 20 and 40.

(f) Persons in state $i0$, $i2$ and $i4$ have not purchased any insurance yet, while persons in state $i1$, $i3$ and $i5$ have purchased insurance. The intensities μ_x^{i01}, μ_x^{i23} and μ_x^{i45} are the annualized purchase rates. The meaning of the purchase rate is two-fold. Firstly, it represents the size of the market. A purchase rate of 0.05 per annum represents what we call a large market. A purchase rate of 0.01 per annum represents what we call a small market. These are assumed to apply in sub-population 1, in which no question of genetic information arises. We call these the 'normal' purchase rates. In sub-populations 2–21, persons who do not have a family history and genetic test

result, buy insurance at the normal rate in all circumstances, because these persons do not have a family history, meaning they will be treated by insurance companies as standard risks. In a large market, persons who have a family history are assumed to have three choices of their purchasing behaviours: normal rate, half of normal rate, or nil, meaning they do not purchase insurance at all. In a small market, these persons are assumed not to purchase insurance at all. Persons who have a family history but have had an adverse test result, are assumed to buy insurance at the normal rate.

(g) A moratorium may be imposed to prevent an insurance company from using genetic information in underwriting. Three types of moratoria are considered here: moratorium on using all genetic test results, moratorium on using adverse genetic test results, and moratorium on using both family history and genetic test results. The effect of a moratorium is to partition the population into different underwriting classes. Persons in each underwriting class are assumed, by the insurers, to carry the same level of risk and will be charged the same rate of premium. The first and second moratoria will partition the population into two underwriting classes, namely family history class and non-family history class. The third moratorium will put all the population into a single underwriting class.

(h) Adverse selection is represented by increasing the insurance purchasing rate of persons with adverse genetic information when a moratorium on all genetic test results or adverse genetic test results is imposed. Moderate adverse selection is represented by a purchasing rate double the normal rate (0.10 per annum in a large market and 0.02 per annum in a small market), and severe adverse selection is represented by a purchasing rate of 0.25 per annum in both large and small markets. A moratorium on genetic test results and family history causes premiums to increase in two different ways.

(a) Persons in the higher-risk sub-population now could purchase the normal amount of insurance premium at ordinary prmeium rates. A new underwriting class forms, comprising persons in all sub-populations. This will increase the premium rate. But this is arguably not adverse selection if their behavior is just the same as that of person in the low-risk sub-populations.

(b) Further, insurance buyers may increase their purchase rate in reaction to the information they have and the relatively lower premiums they have been charged. In this way, adverse selection arises.

Adverse selection could also be represented by increasing the sum assured. It is easy to see that in this case, the premium increases are proportionate to the multiple of the average sum assured taken out by 'adverse selectors', so for brevity we omit the results (Gutiérrez & Macdonald, 2004).

(i) This model is deceptively simple. All states in the dashed box have access to States $i6$ and $i7$. Persons transferring from uninsured states into state $i6$ and $i7$ before they buy insurance become uninsurable. The states, "Onset of relevant diseases", "Other CI events", and "CI payment" made after onset in the cases of MD and HD (See Chapter 2 for details) are concealed in the state "CI". This is for ease of demonstration. Taking the case of MD as an example, we draw a detailed expanded CI market model in Figure B.1 in Appendix B, in contrast to the simplified version in Figure 2.6. Please see Table 4.1 for more detailed information about each relevant disease.

(j) In this market model, insurance applicants can purchase insurance at any age (all policies assumed to expire at age 60), therefore a level premium would be dependent on both entry age and policy term. That is, level premiums are not adapted to the Markov model with age as the time parameter. We need to calculate premium rates

Table 4.1: A table listing the six genetic disorders, the causal mutations and relevant references

Genetic Disorders	Mutations	Section in this thesis	Reference
APKD	APKD1 and APKD2	Section 3.1	Gutiérrez & Macdonald (2003, 2007)
EOAD	PSEN-1	Section 3.2	Gui & Macdonald (2002)
			Espinosa & Macdonald (2007)
HD	Huntingtin gene (HTT)	Section 3.3	Gutiérrez & Macdonald (2004)
MD	DMPK gene	Chapter 2	Chapter 2
HNPCC	MLH1 & MSH2	Section 3.4	Lu *et al.* (2007)
BC & OC	BRCA1 & BRCA2	Section 3.5	Wekwete (2002), Gui *et al.* (2006)

that obey the equivalence principle, but that depend only on age x. The premium rate for underwriting class $\mathcal{C}$ is:

$$\rho^{\mathcal{C}}_{x+t} = \frac{\sum_{ij\in\mathcal{C}} p_i \ {}_tp^{i0j}_x \mu^{ij6}_{x+t}}{\sum_{ij\in\mathcal{C}} p_i \ {}_tp^{i0j}_x}, \tag{4.1}$$

where p_i is the proportion who start in State $i0$ at age x (State $i2$ for sub-populations 2–13), and label ij stands for the state j in the CI market model for sub-population i, has this property and satisfies the equivalence principle.

(k) Using these premium rates, we apply Thiele's equations to calculate the expected insurance loss conditional on being in any state. The premium rates are calculated assuming that there is no adverse selection, in which case the total insurance loss is zero, i.e. EPV of loss without adverse selection $= 0$, because the equivalence

principle is correctly applied. However, if there is adverse selection, the expected insurance loss will be non-zero. This is the cost of adverse selection. To cover the cost, insurance companies will need to increase all rates of insurance premium uniformly by the proportion:

$$\frac{\text{EPV of loss with adverse selection} - \text{EPV of loss without adverse selection}}{\text{EPV of premiums payable with adverse selection}}. \tag{4.2}$$

We take this increase in the premium as our measure of the cost of adverse selection.

4.2 The Cost of Adverse Selection in the Overall Model of the Critical Illness Insurance Market

4.2.1 Moratoria on Using Genetic Test Results

When moratoria on genetic test results are imposed, either on all test results or on negative test results, there are two underwriting classes (UC), that we call UC-I and UC-II. In UC-I, persons will be charged standard premiums and in UC-II, persons will be charged extra premiums because they have family histories. Table 4.2 displays the cost of adverse selection following a moratorium on the use of genetic test results, with family history underwriting allowed, in a CI insurance market operating between ages 20 and 60. We set the genetic testing rate at 0.014 per annum between ages 20 and 40 as our baseline, and compare it with another two scenarios, 0.014 per annum between ages 20 and 60, and 0.035 per annum between age 20 and 40. Table 4.3 shows the results of severe adverse selection, assuming the moratorium is imposed on using genetic test results but not on using family history, and the CI market operates between ages 20 to 60. In this case, we only assume the genetic testing rate to be 0.035 per annum between age 20 and 40.

Generally, the increases are all very small, even with severe adverse selection, although these increases are certainly larger than that caused by any single genetic disorders on its own. We can observe the following features:

(a) In Table 4.2, the costs of adverse selection are generally not substantial, if we take 0.1% as the threshold. Only in a small market and when the adverse selection is extreme, does the cost reach as high as 0.3%–0.4%.

(b) The costs are more substantial in the smaller market and when there is extreme adverse selection. In Table 4.3, we can see that in a small market insurers might have to increase their premium rates generally by 1% in order to recover the loss.

(c) Premium increases are lower if the moratorium applies only to adverse test results, this is because tested persons who are not mutation carriers will be "exonerated" and charged standard premiums. These persons are removed from the underwriting class rated for family history, which then contains a higher proportion of mutation carriers, so the premium charged in respect of this class is higher.

(d) Longer period of genetic testing has little effect on the cost of adverse selection. This is because of high penetrance of all these mutations. Only very few applicants carrying any of these mutations will remain healthy after age 40.

(e) The cost of adverse selection is greater for females than for males. This is because females are at more risk of cancers than males in the case of HNPCC, and males are not at risk of BC and OC at all.

Table 4.2: Percentage increases in CI insurance premium rates arising from moderate adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. CI market operating between ages 20 and 60.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Females	Males	Females	Males
				%	%	%	%
Large	Normal	0.014	20–40	0.01301	0.01242	0.01210	0.01155
	Half	0.014	20–40	0.03182	0.03037	0.01396	0.01332
	Nil	0.014	20–40	0.15143	0.14531	0.04368	0.04176
Small	Nil	0.014	20–40	0.17318	0.16273	0.07778	0.07304
Large	Normal	0.014	20–60	0.01443	0.01376	0.01339	0.01277
	Half	0.014	20–60	0.03629	0.03455	0.01618	0.01542
	Nil	0.014	20–60	0.17390	0.16715	0.05255	0.05038
Small	Nil	0.014	20–60	0.19230	0.18089	0.08689	0.08170
Large	Normal	0.035	20–40	0.02785	0.02659	0.02311	0.02206
	Half	0.035	20–40	0.06756	0.06441	0.02397	0.02287
	Nil	0.035	20–40	0.32112	0.30797	0.09134	0.08725
Small	Nil	0.035	20–40	0.37104	0.34851	0.16621	0.15603

Table 4.3: Percentage increases in CI insurance premium rates arising from severe adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. CI market operating between ages 20 and 60.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Females	Males	Females	Males
				%	%	%	%
Large	Normal	0.035	20–40	0.04835	0.04623	0.03979	0.03805
	Half	0.035	20–40	0.09299	0.08877	0.04139	0.03955
	Nil	0.035	20–40	0.39975	0.38243	0.16993	0.16168
Small	Nil	0.035	20–40	1.23750	1.15633	1.03234	0.96352

4.2.2 Moratoria on Using Genetic Test Results and Family History

A moratorium on genetic test results and family history causes premiums to increase in two different ways.

(a) Persons in the higher-risk sub-populations now could purchase normal amounts of insurance cover at ordinary rates. A new underwriting class forms, comprising persons in all sub-populations. This will increase the premium rate. But this is arguably not adverse selection if their behaviour is just the same as that of persons in the low-risk sub-populations.

(b) Further, insurance buyers may increase their purchase rate in reaction to the information they have and the relatively lower premiums they have been charged. In this way, adverse selection arises.

Table 4.4 shows the increases in standard premium rates arising from the creation of the new underwriting class, and also those arising from moderate or severe adverse selection when a moratorium on the use of all genetic test results and family history is imposed, assuming the CI market operates between ages 20 and 60. The rate of genetic testing is 0.014 per annum with moderate adverse selection, and 0.035 per annum with severe adverse selection between ages 20 and 40. We can observe that the highest cost of adverse selection (premium increases of over 2%) appears in a small market and when there is severe adverse selection, and in a large market, the cost of adverse selection is much smaller than the case in a small market.

Table 4.4: Percentage increases in standard premium rates for CI insurance arising from new underwriting classes, and in all premiums arising from moderate or severe adverse selection, following a moratorium on the use of all genetic test results and family history. CI market operating between ages 20 and 60.

	OR Premium Increase Arising From New		Premium Increase Arising From		Premium Increase Arising From	
Size of	Underwriting Classes		Moderate Adverse Selection		Severe Adverse Selection	
Market	Females	Males	Females	Males	Females	Males
	%	%	%	%	%	%
Large	0.76114	0.73181	0.18087	0.17420	0.30023	0.28954
Small	0.71512	0.67144	0.42624	0.39955	2.13984	2.01271

4.3 The Cost of Adverse Selection in the Overall Model of the Life Insurance Market

In the previous section, we discussed the overall impact of genetic information on a CI insurance market. In this section, we concentrate on a life insurance market. Figure 4.2 shows a life insurance market model, which we used to quantify the impact of genetic information.

As in the CI insurance market model, we partition the population into 21 risk sub-populations depending on whether they are mutation carriers, have family histories, or have family members carrying mutations. Please see Section 4.1 for details. This model is very similar to the CI insurance model, except for the following:

(a) The state $i6$ $i7$ and $i8$, "Onset of Relevant Diseases", stands for the onset of all diseases we discussed, including APKD, EOAD, HD, MD, HNPCC and BC & OC.

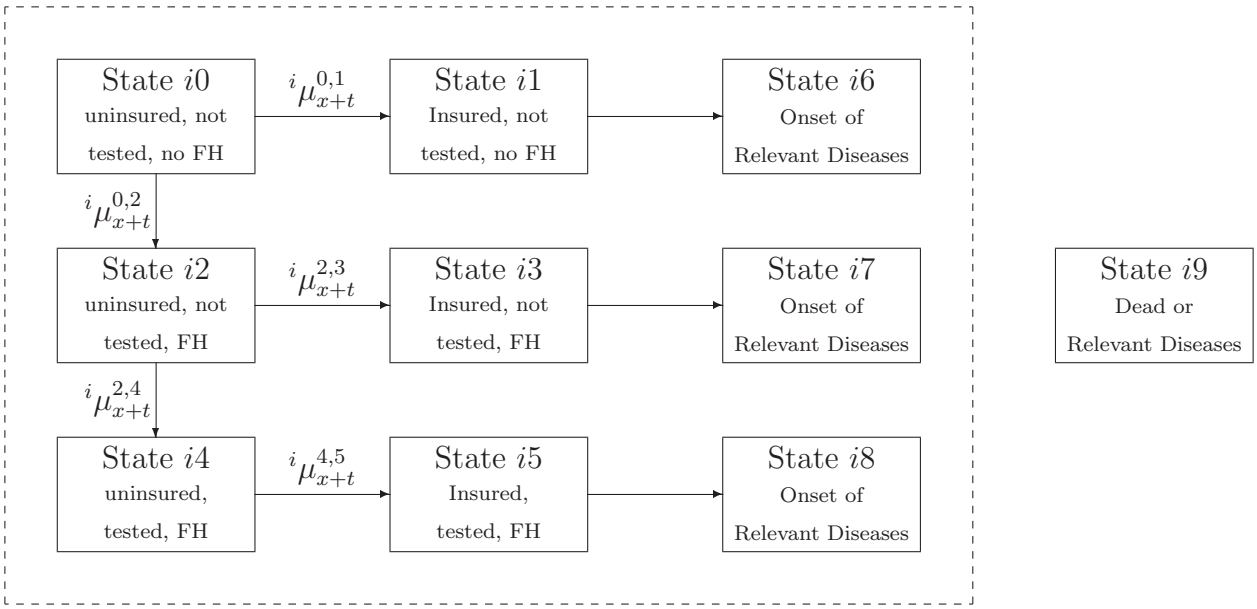

Figure 4.2: A semi-Markov model of family history, genetic testing, insurance purchase and life insurance events for a person in the i^{th} risk sub-population (FH = family history present).

Please refer to Part II for more information about the onset rates of these diseases.

(b) All states in the dashed box have the access into state $i9$, "Dead or Relevant Diseases". Persons transferring from uninsured states into state $i9$ before they buy insurance become uninsurable.

(c) The intensity of entry into state $i9$ from the uninsured state is the sum of the relevant onset rates and the rate of mortality. The intensities of entry into "Onset of Relevant Diseases" states from the insured states are the relevant onset rates. The intensities of entry into state $i9$ after onset of relevant diseases are the post-onset mortality rates. Bear in mind that the post-onset mortality rates depend on both age and duration, so this model is semi-Markov.

(d) As in CI insurance market, we need to charge a premium which depends only on age x according to the equivalent principle. The premium rate for underwriting class $\mathcal{C}$ is:

$$\rho^{\mathcal{C}}_{x+t} = \frac{\sum_{ij\in\mathcal{C}} p_i \left({}_tp^{i0j}_x \mu^{ij9}_{x+t} + \int_0^t {}_{t,z}p^{i0j}_x \mu^{ij9}_{x+t,z} dz \right)}{\sum_{ij\in\mathcal{C}} p_i \left({}_tp^{i0j}_x + \int_0^t {}_{t,z}p^{i0j}_x dz \right)}, \tag{4.3}$$

where label ij stands for state j shown in Figure 4.2 for sub-population i.

(e) We use the trick we introduced in Part I to bring the semi-Markov model to a Markov model regime, by 'paying' to (a reinsurer, presumably) a premium equal to the policy value on entering an "Onset of Relevant Diseases" state from an insured state.

Other features of this model, e.g. the purchase rates, genetic test rates, the onset rates of a family history, the various moratoria, adverse selection, and the measure of

the cost of adverse selection, were all fully introduced in the context of the CI insurance market model. Please see Section 4.1 for more details.

4.3.1 Moratoria on Genetic Test Results

Table 4.5 shows the costs of adverse selection in a life insurance market, assuming a moratorium is imposed on using genetic test results but not on family history, and the life insurance market operates between ages 20 to 60. We set the genetic testing rate at 0.014 per annum between ages 20 and 40 as our baseline, and compare it with another two scenarios, 0.014 per annum between ages 20 and 60, and 0.035 per annum between age 20 and 40. Table 4.6 shows the results of severe adverse selection, assuming a moratorium is imposed on using genetic test results but not on family history, and the life insurance market operates between ages 20 to 60. In Table 4.6, we only assume the genetic testing rate to be 0.035 per annum. Generally, the increases in life insurance premiums are small, even with severe adverse selection. We can observe the following features:

(a) In Table 4.5, the costs of adverse selection are generally not substantial, if we take 0.1% as the threshold. Only in the small market could the cost be high, e.g. between 0.1% and 0.16%.

(b) The costs of adverse selection are more substantial in the smaller market and when there is extreme adverse selection. In Table 4.6, we can see that in a small market insurers have to increase their premium rates generally by nearly 1% in order to recover the loss, which cannot be treated as negligible.

(c) Premium increases are lower if the moratorium applies only to adverse test results, this is because tested persons who are not mutation carriers will be "exonerated" and charged standard premiums. These people are therefore removed from the

underwriting class rated for family history, which then contains a higher proportion of mutation carriers, so the premium charged in respect of this class is higher.

(d) Longer periods of genetic testing have little effect on the cost of adverse selection. This is because of high penetrance of genetic disorders. Only very few applicants carrying mutations will remain healthy after age 40.

(e) The cost of adverse selection is greater in females than in males. This is because females are at more risks of cancers than males in the case of HNPCC, and especially males are not at risk of BC and OC at all.

(f) Noticing that CI insurance and life insurance are both protection type, the magnitude of the cost of adverse selection largely depends on the difference between the premium rates for UC-I and UC-II. In the case of MD, we can see that the extra premium expressed as percentages of standard risk is higher in respect of CI insurance than that in respect of life insurance. Therefore the cost of adverse selection is lighter in a life insurance market than in a CI insurance market.

Table 4.5: Percentage increases in life insurance premium rates arising from moderate adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. Life insurance market operating between ages 20 and 60.

Size of Market	Insurance Purchase of At-Risk Individuals	Rate of Testing	Age Range of Testing	Moratorium on Using All test results		Moratorium on Using Adverse test results	
				Females	Males	Females	Males
				%	%	%	%
Large	Normal	0.014	20–40	0.01281	0.00832	0.01274	0.00830
	Half	0.014	20–40	0.02895	0.01900	0.01534	0.01001
	Nil	0.014	20–40	0.07217	0.05170	0.02217	0.01568
Small	Nil	0.014	20–40	0.07712	0.05633	0.03499	0.02549
Large	Normal	0.014	20–60	0.01348	0.00883	0.01341	0.00809
	Half	0.014	20–60	0.03103	0.02056	0.01658	0.01095
	Nil	0.014	20–60	0.07968	0.05754	0.02518	0.01801
Small	Nil	0.014	20–60	0.08327	0.06121	0.03793	0.02782
Large	Normal	0.035	20–40	0.02800	0.01816	0.02764	0.01803
	Half	0.035	20–40	0.06272	0.04106	0.03274	0.02141
	Nil	0.035	20–40	0.15418	0.11039	0.04680	0.03310
Small	Nil	0.035	20–40	0.16631	0.12138	0.07529	0.05481

Table 4.6: Percentage increases in life insurance premium rates arising from severe adverse selection. Moratoria on the use of genetic test results, family history underwriting still allowed. Life insurancce market operating between ages 20 and 60.

	Insurance Purchase		Age	Moratorium on Using			
Size of	of At-Risk	Rate of	Range of	All test results		Adverse test results	
Market	Individuals	Testing	Testing	Females	Males	Females	Males
				%	%	%	%
Large	Normal	0.035	20–40	0.05564	0.03553	0.05490	0.03523
	Half	0.035	20–40	0.09622	0.06213	0.06558	0.04218
	Nil	0.035	20–40	0.19926	0.01420	0.09186	0.06472
Small	Nil	0.035	20–40	0.60482	0.43511	0.51366	0.36842

4.3.2 Moratoria on Using Genetic Test Results and Family History

As introduced in Section 4.2.2, premium rates increase for two reasons when a moratorium is imposed on using genetic test results and family history. Table 4.7 shows the increases in standard premium rates arising from new underwriting classes and also in all premium arising from moderate or severe adverse selection, when a moratorium on the use of all genetic test results and family history is imposed, assuming the life insurance market operates between ages 20 and 60. The rate of genetic testing is 0.014 per annum with moderate adverse selection, and 0.035 per annum with severe adverse selection between ages 20 and 40. We can observe that:

(a) The highest cost of adverse selection, over 1% of premium, appears in a small market and when there is severe adverse selection. In a large market, the cost of adverse selection, although not negligible, is still much smaller than the case in a small market.

(b) The cost of adverse selection in a life insurance market is lighter that the corresponding case in a CI market.

4.4 Conclusions

In Chapters 2 and 3, we reviewed the epidemiology of our selected genetic disorders. In this chapter, we brought all these together and constructed a semi-Markov model to assess the impact of genetic information relating to all these diseases on the CI or life insurance market. Based on the quantitative results presented in Sections 4.2 and 4.3, we summarise as follows:

Table 4.7: Percentage increases in standard premium rates for CI insurance arising from new underwriting classes, and in all premiums arising from moderate or severe adverse selection, following a moratorium on the use of all genetic test results and family history. Life insurance market operating between ages 20 and 60.

	OR Premium Increase Arising From New		Premium Increase Arising From		Premium Increase Arising From	
Size of	Underwriting Classes		Moderate Adverse Selection		Severe Adverse Selection	
Market	Females	Males	Females	Males	Females	Males
	%	%	%	%	%	%
Large	0.82454	0.51648	0.11460	0.06955	0.20705	0.12547
Small	0.72673	0.44995	0.21477	0.13170	1.26950	0.77738

(a) Under a moratorium on using genetic testing results, with moderate adverse selection, the cost of adverse selection is not substantial, if we take 0.1% of total premium as the threshold, with few exceptions which appear in the case of the small market. With severe adverse selection, the costs of adverse selection become more noticeable in the small market.

(b) A moratorium on using both genetic test results and family history will increase the premium in two ways: consolidation of underwriting classes and adverse selection. The premium increases due to the former are noticeable in both large and small markets. The premium increases arising from the latter are also noticeable. The highest premium increases are about 2% in the CI insurance market and 1% in the life insurance market, in both cases for small markets and extreme adverse selection with a high rate of genetic testing (0.035 per annum).

(c) In both the CI insurance market and the life insurance market, insurers' concerns

have been that the premium increases could lead to market collapse. This is speculative and an economic model for insurance demand would be needed to study whether such an 'adverse selection spiral' (see Section 1.2.3) would result, or whether the market would reach a new equilibrium. Macdonald & Tapadar (2010) looked at this question and concluded that 'no convincing evidence that adverse selection is a serious insurance risk, even if information about multifactorial genetic disorders remains private'. Therefore, we shall believe that an 'adverse selection spiral' is speculative and should not cause significant costs for the insurance companies.

(d) In the "bottom-up" approach, we selected six genetic disorders to study. These disorders are chosen because of their representative impact on the insurance industry. We should ask a question: whether more genetic disorders should be included in the "bottom-up" approach? We answer this question as follows:

(1) In Section 1.2.5, we mentioned that ABI made a list of eight genetic disorders. We took the ABI list as our starting point, and with some adjustments, eventually ended with our selection of six genetic disorders. Please see Section 1.2.5 for detail of these selection.

(2) Among these six genetic disorders, we chose 4 typical autosomal dominant single-gene disorders, namely APKD, EOAD, HD and MD, generally with age-at-onset between age 20 and 60, so that people at risk of these disorders have significant impact on insurance. We also chose two common disorders, namely HNPCC and BC & OC, which could be either hereditary or sporadic. These two disorders are among the most common diseases in the world: CRC is the third most common form of cancer and second leading cause of cancer-related death; BC is the most common cancer and OC is the second common cancer in women. Therefore, we believe that our selection has considered most genetic disorders which potentially have significant impact on the insurance industry.

(3) As mentioned in Section 1.2.5, several genetic disorders have not been modelled mathematically, e.g. MEN and FAP, because these diseases are typical examples of cancer for which genetic testing and early treatment should lead to substantially better outcomes. Inclusion of these genetic disorders will result in lighter cost of adverse selection.

Therefore, we shall trust that our selection of genetic disorders is adequate and our results are robust.

(e) We constructed Markov or semi-Markov models in this chapter to assess the impact of genetic information on the insurance industry, applying the epidemiology relating to each disease. The onset rates of these diseases, because of the retrospective ascertainment scheme, are generally over-estimated, which results in overestimated potential cost of adverse selection. However, we find in Sections 4.2 and 4.3 that in most cases the cost of adverse selection is not noticeable, generally because of the low prevalence of genetic disorders.

(f) Quantification of the impact of genetic information on the insurance industry has been approached from different perspectives, including the "top-down" and "bottom-up" approaches. The purposes of these approaches are not to pinpoint the exact cost of genetic risk, which is very hard, but to set a bound for genetic risk, so that we shall know the magnitude of genetic risk to insurers. Macdonald (2003b) concluded that the increase in life insurance premium rates should be under 10% with a moratorium on using both genetic testing results and family history. That paper used the "top-down" approach with very broad assumptions about the genetic morbidity. In this thesis, we narrowed down this range further to 1% in respect of life insurance. Therefore, we conclude that our results are consistent with previous results and should be a reliable upper bound for measuring the genetic risk.

(g) As in (b), we concluded that the genetic risk might give rise to a potential cost of adverse selection: 2% in the CI insurance market and 1% in the life insurance market. However, the genetic risk has been overwhelmed by the development of health care and general improvement of life expectancy. We shall see in the next Chapter the positive effect of CRC Screening Programs on life insurance premium rates.

Chapter 5

Colorectal Cancer (CRC) and CRC Screening Programs

Colorectal cancer (CRC) is the third most common form of cancer and the second leading cause of cancer-related death in the western world. In Section 3.4, we detailed some features of CRC. CRC can take many years to develop and early detection of CRC greatly improves the chance of survival. Therefore, screening for the disease is recommended for individuals who have been identified as being at risk of CRC. In this chapter, we discuss the impact of a CRC screening program on the onset rate of CRC, mortality caused by CRC, and life insurance premiums.

We will describe examples of two types of screening programs. In one, the entire population is screened. A pilot study for such a program has been undertaken in the UK, called the Bowel Screening Program (BCSP). We describe this, and supporting studies, in Sections 5.1 and 5.2. In the other, members of families known to be at risk of HNPCC are screened. An example, that we will use in our application, is the CRC Surveillance Program in Finland. We describe this, and related work, in Section 5.3 and 5.4.

5.1 The Bowel Screening Program in the UK

The development of the BCSP started with the UK Colorectal Cancer Screening Pilot ('the Pilot'), in order to determine the feasibility of screening for CRC in the UK population using faecal occult blood (FOB) testing. A key task of the Pilot has been to determine whether outcomes achieved in the trial settings can be repeated in population-based programmes. The Pilot commissioned two sites (one in central England, the other in Scotland). Screening began at the Scottish site on March 31, 2000 and at the English site on September 6, 2000. Up to February 2003, 486,355 people had been offered screening at the two sites. The age range of the sampled population is 50 to 70 for both males and females. The final report on the UK Pilot ('the report') was made available online at `http://www.cancerscreening.nhs.uk/bowel/finalreport.pdf`. Some important conclusions were drawn in the report as follows:

(a) The Pilot achieved uptake of the FOB test of close to the target of 60%. However some sub-groups in the population showed lower uptake, including men, younger people, those from deprived areas and individuals of ethnic origin. Uptake was slight less in Scotland than in England.

(b) Staging distribution data indicated a stage-shift towards less advanced cancer, which proved that the screening program helps to detect CRC at an early stage.

The report concluded that benefits had been observed in the Pilot, including CRC-specific mortality reductions, and it should be extended nationally, as the NHS Bowel Screening Program (BCSP). On the official website of the NHS BCSP, we can find the following features:

(a) The BCSP offers routine screening every one or two years to all men and women aged 60–69.

(b) In a faceal occult blood (FOB) test, a small sample of faeces is smeared onto a piece of card. A chemical test detects if any blood is present, at a much lower level than would be evident to the subjects.

(c) Around 98 in 100 people will receive a normal result and will be returned to routine screening. Around 2 in 100 people will receive an abnormal result. They will be referred for further investigation and usually offered a colonoscopy.

(d) A colonoscopy is an investigation that involves looking directly at the lining of the large bowel, using a thin, flexible tube and a tiny camera. If polyps are found, most can be removed painlessly, using a wire loop passed down the colonoscope tube. These tissue samples are then checked for any abnormal cells that might be cancerous. Around 5 in 10 people who have a colonoscopy will have a normal result. About 4 in 10 will be found to have a polyp, which if removed may prevent cancer developing. About 1 in 10 people will be found to have cancer when they have a colonoscopy.

5.2 Additional Evidence for the Effect of Bowel Screening Programs

The final report on the UK Pilot study concluded that it had reduced the mortality caused by CRC. We review briefly other papers which have studied this question.

(a) Mandel *et al.* (1993)

In this study, 46,551 participants aged 50 to 80 years were randomly assigned to screening for CRC once a year, to screening every two years, or to a control group. Participants who tested positive underwent a diagnostic evaluation that included colonoscopy. There was 13 years of follow-up, and causes of death were determined

by a committee. The 13-year cumulative mortality from CRC was 5.88 per 1000 in the annually screened group, 8.33 in the biennially screened group, and 8.83 in the control group. The paper concluded that the mortaliy in the annually screened group, but not the biennially screened group, was significantly lower than in the control group.

(b) Hardcastle *et al.* (1996)

This study recruited 152,850 participants aged 45–74 years who lived in the Nottingham area of the UK between February 1981 and January 1991. The participants were randomly allocated to FOB screening (study group; 76,466) or no screening (control group; 76,384). 2,599 of the participants could not be traced or had emigrated and were excluded from the analysis, leaving 75,253 participants in the screening group and 74,998 in the control group. Control group participants were not told about the study and received no intervention. Study group participants were sent a FOB test kit with instructions from their family doctors. Individuals tested negative were retested every 2 years after the first test. Individuals tested positive were offered colonoscopy. All participants were followed up until June 1995. The median follow-up time was 7.8 years. During the study, 236 cases of CRC were detected in the study group, and 856 in the control group. The distribution of stages of CRC favored the study group, that is, the proportion of CRC cases in early stage was higher in the study group than that in the control group. 360 people died from CRC in the study group compared with 420 in the control group – a 15% reduction in CRC mortality in the study group. The paper concluded that the FOB test should be introduced through a national programme, in order to reduce CRC mortality in the general population.

(c) Kronborg *et al.* (1996)

This study recruited 137,485 people, aged 45–75 years, living in Funen, Denmark, in August 1985, of which, 3 per 14 were randomly allocated to the screening group (study group; 30,967), another 3 per 14 to the non-screening group (control group; 30,966), and the others were not enrolled. The study group participants were screened every 2 years, resulting in 5 rounds of screening during a 10-year follow-up. Participants tested positive were offered full examination and colonoscopy if necessary. In the study group, 481 people had a diagnosis of CRC and 483 in the control group. There were 205 deaths caused by CRC in the study group, compared with 249 in the control group – a 18% reduction in CRC mortality. It was concluded in the study that biennial screening by FOB tests can reduce CRC mortality.

A recent meta-analysis, Hewitson *et al.* (2007), identified 9 articles (including all the above studies), totalling 320,000 participants with follow-up ranging from 8 to 18 years. Some important conclusions are listed as follows:

(a) Combined results show that participants allocated to screening had a 16% reduction in the relative risk (RR) of CRC mortality (RR 0.84, CI: 0.78-0.90).

(b) In 3 studies that used biennial screening, there was a 15% relative risk reduction (RR 0.85, CI: 0.78-0.92) in CRC mortality.

(c) The ages of the sampled population in the above studies range from 45 onwards.

These results are more reliable than each individual study included, because of the much larger sample. Therefore, we will use these results in our study of the effect of reduced CRC mortality on life insurance.

5.3 The CRC Surveillance Program

The CRC Surveillance Program specifically concentrates on persons at risk of HNPCC in Finland. Järvinen (2006) gave a good introduction.We summarise it as follows:

(a) The CRC Surveillance Program begins with ascertainment of HNPCC families by careful inquiry about the family history of cancers in new patients (index patients) with colorectal cancer (CRC) or endometrial cancer (EC). The Amsterdam Criteria, or other rules, might be used to ascertain the HNPCC families. Definite diagnosis of HNPCC requires identification of mutations in one of the mismatch repair genes (see Section 3.4 for details). If found, genetic testing of the first-degree relatives (FDRs) of the index patients will help identify mutation-positive persons predisposed to cancers. Then the detected mutation carriers will go through a cancer prevention program, in order to find and treat colorectal adenomas by colonoscopy before cancers develop.

(b) For ascertained HNPCC family members, the optimal surveillance interval lies between 1 to 3 years beginning from the age of 25. Bliss & Schroy (2004) also mentioned that mutation carriers should undergo surveillance every 1 to 2 years at age 20 years and then annually after age 40.

(c) The CRC Surveillance Program greatly reduces the risk of cancers for mutation carriers. Other papers, e.g. Renkonen-Sinisalo *et al.* (2000), Bliss & Schroy (2004), Vasen *et al.* (1993) and de Vos tot Nederveen Cappel *et al.* (2002), also drew the same conclusion.

5.4 The Effect of CRC Surveillance Program on the Onset Rate of CRC

A series of papers has been published, in order to test the hypothesis that the CRC Surveillance Program can reduce the onset rate of CRC in persons at risk of HNPCC. We review these briefly.

(a) Vasen *et al.* (1993)

In this paper, the authors reviewed the surveillance procedure, the age at diagnosis of CRC, and the occurrence of CRC. A total of 165 families were sampled from 9 centres in Denmark, Finland, Italy, Japan, the Netherlands, Switzerland and the United States. The authors found that half of the centres advise colonoscopy as the only procedure, the interval between the consecutive examinations varying from 1 year to 3 years. A total of 682 high-risk individuals were followed up for 10 years. Only six cases of CRC were encountered, and these were detected at an early stage. This suggested that the surveillance procedures were effective.

(b) Renkonen-Sinisalo *et al.* (2000)

A total of 150 CRC cases detected in 57 HNPCC families over the last 15 years were assigned either to a surveillance group (35) or a non-surveillance group (115). During the follow-up, the stage distribution of the tumours in the surveillance group (Dukes's A, 50%; B, 35%; C, 16%; D, 0%) was significantly more favorable than that in the non-surveillance group (Dukes' A, 17%; B, 50%; C, 16%; D, 17%). This paper concluded that colonoscopic surveillance enables early detection of CRC. However, participants did not take genetic tests to confirm their mutation-carrying status.

(c) de Vos tot Nederveen Cappel *et al.* (2002)

A total of 114 families registered in the Dutch HNPCC registry, which had a mismatch repair (MMR) gene defect and/or met the clinical criteria for HNPCC, were included in this study. These participants were categorized into three groups, namely group I, II and III. Group I was the surveillance group including all unaffected participants. Group II was the "partially resected colon group". Group III was the "total colectomy group". In all three groups, genetic tests were carried out to ascertain mutation carriers. The cumulative risk of developing CRC for proven mutation carriers in 10 years was 10.5% and 15.7% for groups I and II respectively. The results showed that regular HNPCC surveillance can reduce the onset rate of CRC.

Another two important papers are Järvinen, Mecklin & Sistonen (1995) and Järvinen *et al.* (2000). In Järvinen, Mecklin & Sistonen (1995), 251 asymptomatic individuals, aged 20–66 years, belonging to 22 HNPCC families, were ascertained according to the Amsterdam criteria. The observation on these participants (with one individual added later) continued to the publication of the subsequent paper Järvinen *et al.* (2000). The studies both aimed to evaluate the efficacy of long-term surveillance program by means of colonoscopy and polypectomies. Of the 252 (251 + 1) participants, 133 chose to take colonic examination between 1982 and 1986 (study group), whereas the 119 control participants either declined screening (78) or could not be traced (41). Genetic tests for the mutation segregating in each particular family was offered for 205 individuals and performed in 193 between 1996 and 1998: 116 in the study group and 77 in the controls. Another 9 participants were classified as mutation-positive carriers without a genetic test. In total, there were 44 mutation-positive and 74 mutation-negative individuals in the study group compared with 46 positive and 38 negative in the controls. The examination was repeated at 3-year intervals, and the use of colonoscopy reached nearly 100%. Järvinen *et al.* (2000) concluded that in mutation-positive subjects alone, the onset rate of CRC was reduced by 56%. Unlike the BCSP, which starts from age above 45, the CRC Surveillance

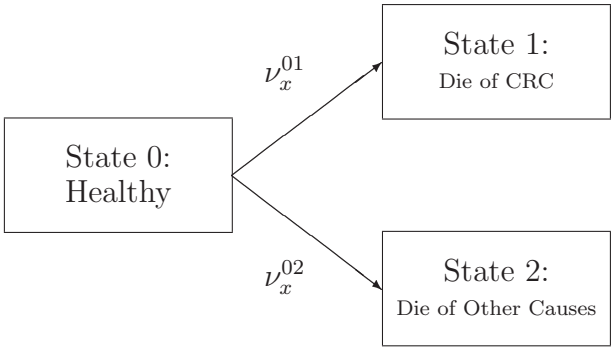

Figure 5.1: A Markov model of life history, differentiating mortality caused by CRC and other mortality.

Program starts from age 20. Therefore, we assume the reduced CRC onset rate takes effect from age 20. Compared with the other three papers listed above, Järvinen *et al.* (2000) included more patients in the study, had longer follow-up and carried out genetic tests to differentiate mutation carriers. Therefore, we choose to use the results presented in Järvinen *et al.* (2000) to evaluate the effect of a surveillance program on life insurance.

5.5 The Effect of a Bowel Screening Program on Life Insurance

We present a Markov model for pricing a life insurance policy in Figure 5.1, in order to evaluate the effect of a Bowel Screening Program on life insurance.

This model splits the total population force of mortality into the force of mortality caused by CRC, ν_x^{01}, and that by other causes ν_x^{02}. We suppose that ELT15 represents the population mortality. Hence

$$\mu_x^{ELT15} = \nu_x^{01} + \nu_x^{02} \tag{5.1}$$

ELT15 was produced using O.N.S. data during 1990–1992 (O.N.S., 1997 & 1999). We also use O.N.S. (1997 & 1999) to estimate the intensity ν_x^{01}. Assume that θ_x^{ELT15} and E_x^{ELT15} are the numbers of deaths and the number exposed to risk at age x during 1990–1992. We can split the total number of deaths θ_x^{ELT15} into the numbers of deaths caused by CRC, denoted $\theta_x^{ELT15,CRC}$, and by all other causes, denoted $\theta_x^{ELT15,Other}$. Hence the crude rate of the mortality caused by CRC, denoted $\dot{\nu}_x^{01}$, can be estimated as:

$$\dot{\nu}_x^{01} = \frac{\theta_x^{ELT15,CRC}}{E_x^{ELT15}}. \tag{5.2}$$

The crude rate $\dot{\nu}_x^{01}$ is then graduated to give our smoothed estimate ν_x^{01}. The total number of deaths caused by CRC, $\theta_x^{ELT15,CRC}$, was found from O.N.S. (1999), a CD-ROM containing the coded death records for the UK during year 1971–1997. We used polynomial functions to fit the crude estimates as follows:

for males:

$$\begin{aligned} \nu_x^{01} &= 5.943 \times 10^{-5} - 2.188 \times 10^{-5} x + 2.404 \times 10^{-6} x^2 \\ &\quad -9.798 \times 10^{-8} x^3 + 1.576 \times 10^{-9} x^4 - 7.078 \times 10^{-12} x^5, \end{aligned} \tag{5.3}$$

and for females:

$$\begin{aligned} \nu_x^{01} &= -1.108592 \times 10^{-5} + 5.639 \times 10^{-6} x - 2.152 \times 10^{-7} x^2 \\ &\quad -7.249 \times 10^{-9} x^3 + 3.589 \times 10^{-10} x^4 - 2.026 \times 10^{-12} x^5. \end{aligned} \tag{5.4}$$

The crude estimates and fitted functions are displayed in Figure 5.2.

Then the intensity ν_x^{02} is easily derived by subtracting the intensity ν_x^{01} from μ_x^{ELT15}. Based on Hewitson *et al.* (2007), we assume that the intensity ν_x^{01} is reduced by about 15% if people go through either annual or biennial screening, and the reduction takes effect

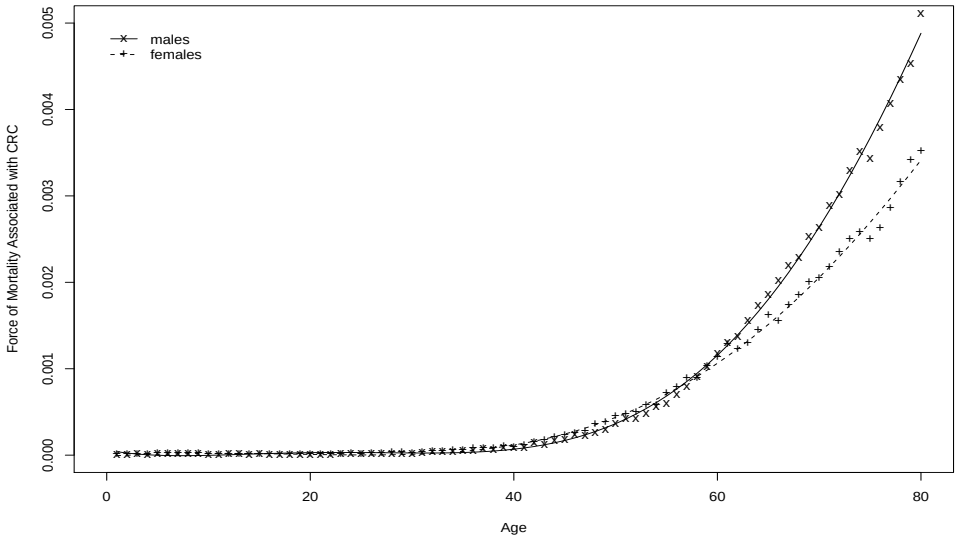

Figure 5.2: Observed and fitted age-related mortality caused by CRC for males and females.

from age 45 onward. We treat the persons who choose not to take part in a screening program as standard risk.

In this model, insured applicants pay premiums continuously while remaining in State 0, and the claim is paid when they enter State 1 or 2. We assume a constant force of interest $\delta = 0.05$. We calculate the net level rate of premium P for a unit sum assured as follows.

$$P \times \int_x^{x+n} e^{-\delta t}{}_t p_x^{00} dt = \int_x^{x+n} e^{-\delta t}{}_t p_x^{00} (\nu_{x+t}^{01} + \nu_{x+t}^{02}) dt. \tag{5.5}$$

We calculate the premium rates for different entry ages and terms for both males and females. Table 5.1 shows the level net life insurance premium rate per unit sum assured for persons not taking part in a screening program, taking part in an annual or biennial screening program. Table 5.1 shows the premium rates as a percentage of the premium

rates for a standard risk.

We can see some features as follows:

(a) There is a fall of 1–2% in premium rates in full term policies expiring at age 60, because CRC risk is greater at high ages than at low ages. In the case of entry age 50 and policy term 10 year, for males, we observe a 2% reduction in premium rates.

(b) The reduction seems greater for males than for females, probably because male CRC mortality is greater than female before screening program, especially at high ages (see Figure 5.2).

5.6 The Effect of a CRC Surveillance Program on Life Insurance

A semi-Markov model for HNPCC in life insurance is presented in Figure 5.3, for a sub-population labelled i defined by genotype.

We stratify the whole population into the following three sub-populations:

(a) sub-population 1: non-mutation carriers;

(b) sub-population 2: mutation MLH1 carriers;

(c) sub-population 3: mutation MSH2 carriers;

We take non-mutation carriers not going through CRC Surveillance Program as standard risks. The onset intensities of CRC, EC (females only) and OECC associated with HNPCC and the post-onset mortalities were introduced in Section 3.4. The onset rates of CRC for mutation carriers going through CRC Surveillance Program will be reduced by 56% (see Section 5.5).

Table 5.1: Level net life insurance premium rates for persons not taking part in a screening program and taking part in a screening program.

		No Screening Program				Screening Program			
Age	Term	Females		Males		Females		Males	
20	10	0.000342	(100%)	0.000867	(100%)	0.000342	(100%)	0.000867	(100%)
	20	0.000459	(100%)	0.000973	(100%)	0.000459	(100%)	0.000973	(100%)
	30	0.000684	(100%)	0.001268	(100%)	0.000679	(99%)	0.001265	(100%)
	40	0.001066	(100%)	0.001874	(100%)	0.001052	(99%)	0.001861	(99%)
30	10	0.000653	(100%)	0.001148	(100%)	0.000653	(100%)	0.001148	(100%)
	20	0.001038	(100%)	0.001687	(100%)	0.001029	(99%)	0.001680	(100%)
	30	0.001681	(100%)	0.002737	(100%)	0.001654	(98%)	0.002713	(99%)
40	10	0.001679	(100%)	0.002592	(100%)	0.001656	(99%)	0.002574	(99%)
	20	0.002759	(100%)	0.004424	(100%)	0.002704	(98%)	0.004375	(99%)
50	10	0.004590	(100%)	0.007589	(100%)	0.004480	(98%)	0.007482	(99%)

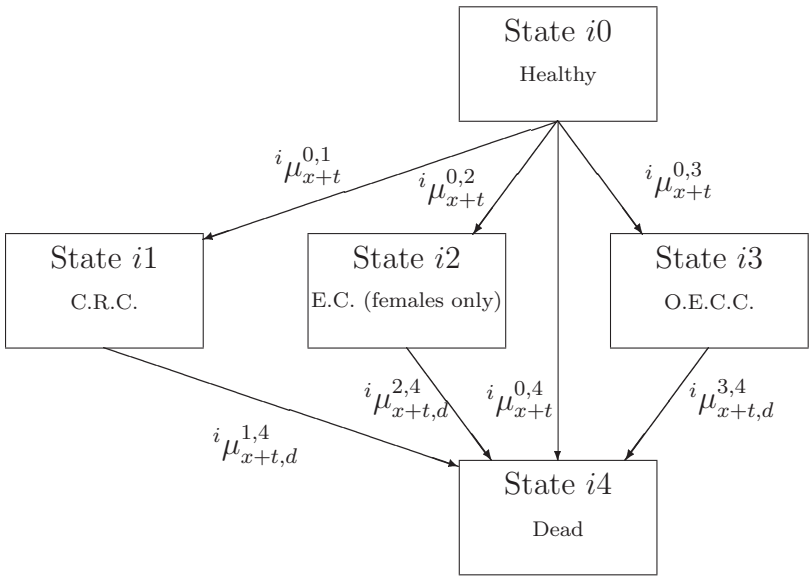

Figure 5.3: A semi-Markov model for HNPCC in life insurance for sub-population i, defined by genotype.

In this model, insured applicants pay premiums continuously while remaining in states $i0$, $i1$, $i2$ or $i3$, and the claim is paid when they enter state $i4$. We calculate the net level premium for a unit sum assured. We assume constant force of interest $\delta = 0.05$. The expressions for the EPVs of a unit of premium and a unit of benefit follow, assuming the entry age is x and policy term is n:

$$\text{EPV[Premium]} = \int_0^n e^{-\delta t} \cdot {}_t^i p_x^{\overline{00}} dt + \sum_{j=1,2,3} \int_0^n e^{-\delta t} \cdot {}_t^i p_x^{\overline{00}} \cdot {}^i\mu_{x+t}^{0j} \int_0^{n-t} e^{-\delta s} \cdot {}_s^i p_{x+t}^{\overline{jj}} \, ds \, dt \quad (5.6)$$

$$\text{EPV[Benefit]} = \int_0^n e^{-\delta t} \cdot {}_t^i p_x^{\overline{00}} \cdot {}^i\mu_{x+t}^{04} dt + \sum_{j=1,2,3} \int_0^n e^{-\delta t} \cdot {}_t^i p_x^{\overline{00}} \cdot {}^i\mu_{x+t}^{0j} \int_0^{n-t} e^{-\delta s} \cdot {}_s^i p_{x+t}^{\overline{jj}} \cdot {}^i\mu_{x+t+s,s}^{j4} \, ds \, dt \quad (5.7)$$

Then the net level premium for a unit of benefit payable continuously should be EPV[Benefit]/EPV[Premium]. Table 5.2 and Table 5.3 show, respectively, the level net premium rates per unit sum assured, and the same as a percentage of standard risk, respectively; for MLH1 and MSH2 mutation carriers not taking part in a CRC Surveillance Program and taking part in a CRC Surveillance Program.

We can see some features from the above two tables.

(a) For persons who choose not to take part in a CRC Surveillance Program, mutation-carrying applicants are uninsurable in most cases, except for a few cases at high ages. For persons who choose to take part in a CRC Surveillance Program, we can see substantial decreases in premium, and most cases become insurable at increased premium rates. However, premium rates for female mutation MSH2 carriers are generally above 300% of the premium rates for standard risks. Therefore they are still probably uninsurable.

(b) Female mutation carriers are more likely to be declined than male mutation carriers, because females are also at risk of EC, which is not screened for.

Table 5.2: Level net life insurance premium rates for MLH1 and MSH2 mutation carriers not taking part in a CRC Surveillance Program and taking part in a Surveillance Program.

		No Surveillance Program				Surveillance Program			
		MLH1		MSH2		MLH1		MSH2	
Age	Term	Females	Males	Females	Males	Females	Males	Females	Males
20	10	0.000976	0.001806	0.000755	0.001714	0.000624	0.001277	0.000634	0.001265
	20	0.002298	0.003404	0.001750	0.003447	0.001334	0.002069	0.001340	0.002136
	30	0.003566	0.004846	0.003251	0.005479	0.002180	0.002941	0.002444	0.003356
	40	0.004638	0.006095	0.004523	0.006939	0.003037	0.003918	0.003490	0.004536
30	10	0.003189	0.004358	0.002684	0.004956	0.001884	0.002577	0.001999	0.002889
	20	0.005293	0.006775	0.005350	0.008519	0.003310	0.004047	0.003947	0.005022
	30	0.006927	0.008714	0.007381	0.010775	0.004664	0.005623	0.005628	0.006922
40	10	0.005891	0.007298	0.007281	0.010341	0.004094	0.004726	0.005392	0.006289
	20	0.008700	0.010774	0.010725	0.014231	0.006384	0.007494	0.008221	0.009553
50	10	0.009622	0.012608	0.012094	0.015088	0.007745	0.009968	0.009617	0.011588

Table 5.3: Level net life insurance premium rates for MLH1 and MSH2 mutation carriers not taking part in a CRC Surveillance Program and taking part in a CRC Surveillance Program, expressed as percentages of standard risks.

		No Surveillance Program				Surveillance Program			
		MLH1		MSH2		MLH1		MSH2	
Age	Term	Females	Males	Females	Males	Females	Males	Females	Males
		%	%	%	%	%	%	%	%
20	10	285	208	221	198	182	147	185	146
	20	501	350	381	354	291	213	292	220
	30	521	382	475	432	319	232	357	265
	40	435	325	424	370	285	209	327	242
30	10	488	380	411	432	289	224	306	252
	20	510	402	515	505	319	240	380	298
	30	412	318	439	394	277	205	335	253
40	10	351	282	434	399	244	182	321	243
	20	315	244	389	322	231	169	298	216
50	10	210	166	263	199	169	131	210	153

(c) Table 5.2 and Table 5.3 are based on known genetic test results. However, underwriting based on genetic test results is banned in the UK, while family history (FH) underwriting is more realistic. In this case, the net level premium will be the ratio of the EPV of 1 unit of benefit to the EPV of 1 unit of premium each weighted by the probability

$$\mathbf{P}\left[\text{genotype } g_i \mid \text{healthy at age } x \text{ and with family history}\right],$$

for sub-population i. We should expect that more weight would be given to the non-mutation carrier sub-population. As in (a), Table 5.3 shows that most cases going through the CRC Surveillance Program are insurable at increased premium rates with their genetic testing result known to insurers. Because of 'averaging-out' effect, we expect all applicants with a family history taking part in a CRC Surveillance Program to be accepted by life insurers at increased premium rates (or even at an ordinary rate at high ages).

5.7 Conclusions

We give short conclusions as follows:

(a) CRC Screening Programs have been proved to reduce the CRC related risk by early detection and effective treatment. In this chapter, we described two types of CRC Screening Program, one for the entire population, called BCSP in the UK, and the other for HNPCC families, called CRC Surveillance Program, implemented in Finland. In the case of the BCSP, Hewitson *et al.* (2007), a meta-analysis, concluded that participants allocated to screening had a 16% reduction in the relative risk of CRC mortality, whereas in the case of the CRC Surveillance Program, Järvinen *et*

al. (2000) concluded that the onset rate of CRC for mutation carriers was reduced by 56%.

(b) We calculated life insurance premium rates for persons taking part in the BCSP and not taking part in the BCSP. We found that the BCSP may reduce the life insurance premium rates by about 1% (for females aged 50 seeking 10-year policy the reduction in life insurance premium rate is about 2%).

(c) We evaluted life insurance premium rates for persons taking part in the CRC Surveillance Program and those not doing so. We found that the CRC Surveillance Program made mutation MLH1 and MSH2 carriers insurable at increased life insurance premium rates with genetic testing results known to insurers. Hence, under family history underwriting, persons carrying HNPCC family history would also be more likely to be insurable.

(d) We concluded in Chapter 4 that in order to recoup the cost of genetic risk the insurers need to increase the insurance premium rates by about 1% under very extreme circumstances, e.g. severe adverse selection and moratorium on using both genetic testing rests and family history. It is very interesting to see that the BCSP actually appears to negate the genetic risk by about the same magnitude, without even considering the general trend of mortality improvement.

Chapter 6

Income Protection Insurance and Ovarian Cancer

6.1 Income Protection Insurance

An income protection insurance (IPI) policy (formerly known as permanent health insurance) pays the policyholder an income in the event of sickness or disability which prevents them from working. To eliminate trivial or fraudulent claims, there is usually a 'deferred period' (DP) that must elapse after onset or diagnosis of the illness before the claim commences. Usually, the DP is 1, 4, 13, 26 or 52 weeks. The benefit is payable until the earlier of expiry of the policy (usually at age 60 or 65), the policyholder's return to work, or death. Mackay (1993) mentioned some features of IPI policies in the UK.

(a) Unlike life insurance, of which the underwriting is based on mortality risks, the underwriting of IPI considers applicants' morbidity risks.

(b) An IPI policy is non-cancellable by the insurance company. The insurance company does not have the option to cancel the policy, irrespective of the number of claims.

(c) Occupation is a much more important risk factor than for life insurance.

(d) The insurer wishes the insured to have an incentive to return to work. Therefore, most providers limit the maximum that can be insured to 50–60% of gross income.

(e) In general terms, a policyholder receives benefit when the period of disability is longer than the deferred period (DP), and pays premiums when (s)he is healthy or when the period of disability is shorter than the DP. A DP of 52 weeks is not common in the IPI market. Therefore, in this thesis, we consider only DPs of 1, 4, 13 and 26 weeks.

(f) An IPI policy normally expires at age 60 or 65. Note that the current normal retirement age (NRA) is 65 for males, 60 for females. However, NRAs are being equalised in the UK as in some other countries, and it is planned that the NRA for females will rise to 65 or over. In this thesis, we take the age of expiry to be age 65.

Disability is a key definition in an IPI policy. Three main definitions are used in the UK (Mackay, 1993) as follows:

(a) 'Totally unable through sickness or accident to follow one's own occupation and not following any other for profit or reward'

(b) 'Totally unable through sickness or accident to follow one's own occupation or any other to which reasonably suited and not following any other for profit or reward'

(c) During the first two years of disability: 'totally unable through sickness or accident to follow one's own occupation and not following any other for profit or reward'. For disability lasting beyond two years: 'totally unable through sickness or accident to follow any other reasonably suitable occupation and not following any other for profit or reward'.

Despite the slight differences between the above three definitions, they all emphasize that the claimant must be unable to earn income as a result of disability caused by either accident or sickness. However, the definitions are far from exact. The interpretation is left to the claim manager and the policyholder's personal physician. Not surprisingly, disputes can emerge at claim stages.

Varying types of IPI policies have been developed in the UK as follows (Sanders & Silby, 1988):

(a) Individual Conventional Policies

This is a traditional type of IPI policy, providing a regular income during disability. The fundamental feature of the product is that the benefits and premiums are fixed at the start of the policy.

(b) Increasing Benefit Policies

This policy provides some protection against inflation for the policyholders. Other features are similar to those of a conventional policy.

(c) Unit-Linked Policies

The mechanism of a unit-linked IPI policy is very similar to other types of unit-linked policy. The policyholder's premiums after deduction for expenses are invested in unit-linked funds and units are regularly cancelled by the office to pay IPI risk premiums.

In this chapter, our purpose is to investigate the financial impact of genetics on an IPI market. Therefore, for simplicity, we consider only individual conventional policies with level benefits.

6.2 Actuarial Models of IPI

The traditional method that was formerly used in the UK to develop IPI premium rates is known as the Manchester Unity method. In CMIR No.12 (1991), a new method to model an individual IPI policy was proposed and is now widely used. This model, as shown in Figure 6.1, is semi-Markov. The policyholder starts in state 'Healthy' after taking out an IPI policy and may either fall sick or die at any later time, governed by intensities, σ_x and μ_x respectively, where x is age. Once sick, the policyholder may recover or die, governed by intensities $\rho_{x,z}$ and $\nu_{x,z}$, respectively, where x is age and z is the duration of sickness. The intensities $\rho_{x,z}$ and $\nu_{x,z}$ depend on the policyholder's current age x and the disease duration z. In CMIR No.12 (1991), these intensities were graduated based on policy data collected between 1975–1978 and in the later CMIR No.20 (2001), these values were updated using 1995–1998 data.

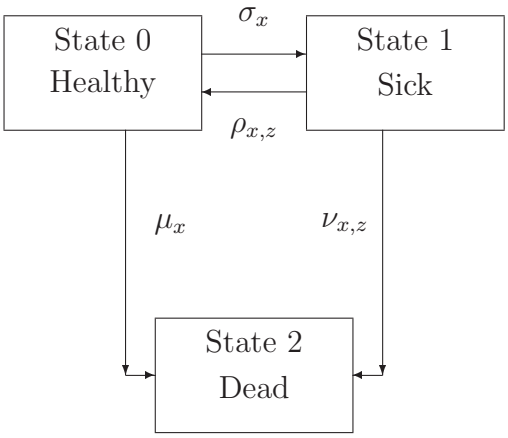

Figure 6.1: The CMIB's three-state semi-Markov model for sickness.

We can use this model, as parametrised by the CMI to calculate the premium rates for standard risk. However, in this study, our purpose is to calculate IPI premium rates for women at risk of ovarian cancer (OC). Unlike models of CI insurance, which need

only to consider onset, or life insurance, which need only to consider survival after onset in addition, we must model onset, progression of the illness and its treatment, recovery, return to work and then possible further episodes. Therefore, we need a model with the following features:

(a) It can can distinguish OC from other sicknesses;

(b) It can differentiate OC diagnosed at different stages;

(c) It can represent treatments of OC depending on the stage;

(d) It includes recoveries following treatment;

(e) It includes recurrence of OC; further treatment and recoveries; and death.

Chapters 6 and 7 follow the same line as Lu, Macdonald & Waters (2008) and Lu *et al.* (2008), which reviewed the epidemiology of BC and constructed semi-Markov models to study the impact of genetic information related to BC in the IPI market. Therefore we next outline the essential features of OC, and then specify a life history model that represents them.

6.3 Ovarian Cancer

In this chapter, our purpose is to model IPI products, which involves the diagnosis of OC at different stages, treatment, recovery and recurrence. Therefore, we give a detailed introduction to the epidemiology of these aspects of OC.

6.3.1 An Introduction to Ovarian Cancer

OC is a malignant tumour on or within an ovary. OC is the fifth leading cause of cancer deaths in women, the leading cause of death from gynaecological malignancy, and the

second most commonly diagnosed gynaecological malignancy among North American and northern European women (DeVita, Hellman & Rosenberg, 1997). Approximately 140,000 new cases of ovarian cancer occur worldwide each year. The highest incidence rates are observed in eastern and northern Europe, North America (range 7.0 – 15.1 per 100,000). The lowest incidence rates occur in northern and western Africa, and Asia, including Japan (range 0.7 – 6.7 per 100,000) (Morrison, Hodgson & Haites, 2002). In the United States, an estimated 22,220 new cases of OC are diagnosed each year and about 16,210 deaths occur annually as a result of OC. The lifetime probability of developing OC in the North American female population is approximately 1.4% (DiSaia & Creasman, 2007). In the UK, almost 7,000 new cases and about 4,500 OC-caused deaths were reported in 1999. The lifetime risk of OC for women in the UK is about 2% (Reznek, 2007).

6.3.2 Classification of Ovarian Cancer

OC can be classified into newly-diagnosed OC and recurrent OC. We briefly introduce these two classes of OC as follows:

Newly-Diagnosed OC

On the basis of distinct clinical and pathologic features, ovarian carcinomas can be separated into three major entities (Auranen & Iselius, 1998):

(a) epithelial carcinomas

(b) germ cell tumours

(c) stromal carcinomas.

The vast majority (85 – 90%) are epithelial in origin. Epithelial ovarian cancer is thought to be derived from the ovarian surface epithelium or from ectopic endometrial or

fallopian tube (tubal) tissue. OC could also form in the egg cells, resulting in a germ cell tumour. These are normally benign and only occur in young women. Stromal carcinomas are a group of tumours of sex cord-derived tissues of the ovary and testis. This group accounts for 8% of OC (Kristensen & Tropé, 1997).

Recurrent OC

Normally the first-line chemotherapy (first-line treatment is for the patients who contract OC for the first time) will be platinum-based. Recurrence of OC happens quite often after first-line treatment. According to whether OC patients respond to the first-line treatment, and the progression-free interval since the finish of the first-line chemotherapy, the recurrence of OC could be classified as: refractory recurrent OC, platinum-resistant recurrent OC and platinum-sensitive recurrent OC (Amos & Struewing, 1993).

Refractory OC means that the patients do not response to the first-line treatment at all. The recurrent OC develops very shortly after the finish of the first-line treatment, or even during the treatment. Platinum-resistant OC means that the patients develop the recurrent OC less than six months after the end of the first-line treatment. Platinum-sensitive OC means that the patients response to the first-line treatment well, and recurrence occurs more than 6 months after finishing the first-line treatment. The difference between refractory OC and platinum-resistant OC is not clear in some epidemiological and medical studies (DiSaia & Creasman, 2007).

6.3.3 Risk Factors

Past studies classified risk factors into three categories: personal characteristics, reproductive factors and environmental factors (Banks, Beral & Reeves, 1997).

(a) Personal Characteristics

Personal characteristics include age, body mass index (BMI) and genetic/familial factors. Age, as a risk factor, is commonly used in insurance underwriting. OC (other than the usually-benign germcell OC) is rare among young women and its incidence increases dramatically with age (Adami *et al.*, 1990). A family history is the single most important risk factor . However, most OCs are sporadic in nature. Less than 5% of cases can be defined as hereditary OC (at least two first degree relatives with OC) in which predisposition for the disease follows a classic pattern of autosomal dominant transmission (Jacobs & Lancaster, 1996). Women with a family history of OC are three to four times more likely to develop OC than those without such a history (Lynch *et al.*, 1978). BRCA1/2 mutations are confirmed to associate with inherited breast cancer and inherited OC (Gayther *et al.*, 1997).

(b) Reproductive Factors

Pregnancy decreases the risk and multiple pregnancies have an increasingly protective effect (Beral, Fraser & Chilvers, 1978). Infertility increases the risk of OC. Drugs that stimulate ovulation such as clomiphene have also been shown to increase risk two- to three-fold when overdosed in infertile women. Users of oral contraceptive pills have a decreased risk (Booth, Beral & Smith, 1989). In a large study by the World Health Organization (WHO), the relative risk for women who ever used oral contraceptives, compared with those who had not, was 0.75 (WHO, 1989). In a study examining the relationship between duration of oral contraceptive use and incidence of epithelial OC, it was demonstrated that 5 years of oral contraceptive use reduced the risk in nulliparous women to the same level as parous women or women who were nonusers (Franceschi *et al.*, 1991). These epidemiologically derived associations between hormonal factors and risk of OC support the "incessant ovulation hypothesis", namely that the risk of OC is the direct function of the number of ovulatory cycles in a woman's life span. The exact nature of the relationship between

menarche, menopause, and OC is not clear (Risch, Marrett & Howe, 1994). The evidence regarding the effect of breast feeding on OC is also inconclusive (Booth, Beral & Smith, 1989).

(c) Environmental Factors

Environmental factors have also been identified to have an influence on the incidence of OC. Industrialised countries, with the exception of Japan, have the highest incidences of ovarian cancer. These observations led to epidemiological studies examining the association between diet and industrial exposure to carcinogens and the subsequent development of OC. However, this is not very conclusive yet (Wehner, 1994). Similarly, coffee and tobacco usage have not been associated with OC although there may be a slight increased risk with alcohol use (Mori, *et al.*, 1988).

6.3.4 Staging of Ovarian Cancer

There are different staging systems in use for OC. A commonly used staging system is the International Federation of Gynaecology and Obstetrics (FIGO) staging system as follows (FIGO, 2000):

(a) In stage I, cancer is found in one or both of the ovaries and has not spread. Stage I is divided into stages IA, IB and IC.

Stage IA: Cancer is found in a single ovary.

Stage IB: Cancer is found in both ovaries.

Stage IC: Cancer is found in one or both ovaries and one of the following is true: Cancer is found on the surface of one or both ovaries; or the capsule (outer covering) of the tumour has ruptured; or cancer cells are found in the fluid of the peritoneal cavity or in washings of the peritoneum.

(b) In stage II, cancer is found in one or both ovaries and has spread into other areas of the pelvis. Stage II is divided into stages IIA, IIB and IIC.

Stage IIA: cancer has spread to the uterus and/or the fallopian tubes.

Stage IIB: cancer has spread to other tissue within the pelvis.

Stage IIC: cancer has spread to the uterus and/or fallopian tubes and/or other tissue within the pelvis and cancer cells are found in the fluid of the peritoneal cavity or in washings of the peritoneum.

(c) In stage III, cancer is found in one or both ovaries and has spread to other parts of the abdomen. Stage III is divided into stages IIIA, IIIB and IIIC.

Stage IIIA: the tumor is found only in the pelvis, but cancer cells have spread to the surface of the peritoneum.

Stage IIIB: cancer has spread to the peritoneum but is 2 centimeters or smaller in diameter.

Stage IIIC: cancer has spread to the peritoneum and is larger than 2 centimeters in diameter and/or has spread to lymph nodes in the abdomen.

(d) In stage IV, cancer is found in one or both ovaries and has metastasised beyond the abdomen to other parts of the body. Cancer is found in the tissues of the liver.

For details of other similar systems, see Högberg, Glimelius & Nygren (2001).

6.3.5 Grading of Cancer

In pathology, grading is a measure of the progress of tumours and other neoplasms. Some pathology grading systems apply only to malignant neoplasms (cancer); others apply also to benign neoplasms. Pathology grading systems are used to classify neoplasms in terms of how abnormal the cells appear microscopically and what may be the outcome in terms

of rate of growth, invasiveness, and dissemination. Cancer is a disorder of excessive cell growth, hence cancer cells often are poorly differentiated. The grade reflects the degree of cellular differentiation and refers to how much the tumor cells resemble or differ from the normal cells of the same tissue type. Grade is a marker of how differentiated a cell is. Grade is rated numerically (Grade 1–4) or descriptively (e.g. "high grade" or"low grade"). The higher the numeric grade, the more "poorly differentiated" is the cell, and it is called "high grade". A low grade cancer has a low number and is "well-differentiated". The following is an example of the grading categories (Benedet, 2000).

(a) GX: Grade cannot be assessed

(b) G1: Well differentiated (Low grade)

(c) G2: Moderately differentiated (Intermediate grade)

(d) G3: Poorly differentiated (High grade)

(e) G4: Undifferentiated (High grade)

The tumor grade, along with the staging, is used to develop an individual treatment plan and to predict the patient's prognosis (Shepherd, 1989).

6.3.6 Diagnosis

Epithelial cancers of the ovary have been described as a "silent killer" because the overwhelming majority of patients present with disease that has spread outside of the ovary and indeed outside of the pelvis at the time of initial presentation. Approximately 70% of patients with epithelial cancers of the ovary present with stage III or IV disease. Abdominal discomfort and bloating are the most common symptoms experienced by women with epithelial ovarian cancer, followed by vaginal bleeding, gastrointestinal symptoms, and

urinary tract symptoms. Patients presenting with non-specific lower abdominal discomfort or bloating require a prompt and careful pelvic examination. The failure to perform routine rectovaginal pelvic examinations may result in women with relatively early-stage ovarian cancer having a delay in diagnosis (Disaia & Creasman, 2007).

6.3.7 Prognosis

Generally, important prognostic factors include staging, grading, peritoneal fluid cytology, etc. Staging is a powerful prognostic predictor in ovarian cancer, and most other putative prognostic factors are of little importance in comparison to staging (Zanetta *et al.*, 1998). Peritoneal fluid cytology is for substaging purpose, e.g. stage IA, IB, etc.

For early stage OC, other factors, e.g. grading, rupture before and during surgery and DNA ploidy might also be very important (Vergote *et al.*, 1993 and 2001). Vergote & Trimbos (2003) concluded that DNA ploidy is the second most important prognostic factor, but this has not been widely accepted. We will omit this from our model.

For late stage OC, chemotherapy, extent of residual disease, and response to platinum-based chemotherapy (see Section 6.3.2 and 6.3.8) are also strong prognostic factors. Shamsunder *et al.* (2000) found that overall survival for patients receiving chemotherapy was significantly higher compared to those who did not. The prognosis is better for patients responding to the first-line platinum-based treatment than those who did not. For platinum-resistant OC patients, prognosis is normally worse after subsequent treatments.

6.3.8 A Brief Introduction to Treatments for OC Patients

Treatment option is an important factor in the outcome of OC. We are interested in the following aspects of treatments:

(a) What are the standard treatments for OC, including newly-diagnosed OC and recurrent OC?

(b) What recovery rates are associated with treatments?

(c) What recurrence rates are associated with recovery after treatment?

Standard treatments for OC are surgery and chemotherapy. Other treatments, e.g. radiotherapy and hormonal therapy, might also be available (Herzog, Holloway & Stuart, 2003).

Surgery

Different types of surgery for OC are as follows:

(a) Total hysterectomy: remove uterus and cervix.

(b) Unilateral salpingo-oophorectomy: remove one ovary and one fallopian tube.

(c) Bilateral salpingo-oophorectomy: remove both ovaries and fallopian tubes.

For OC patients at early stages, the choice of surgery usually depends on whether a women plans to have children. For women who plan to have children, surgery is either unilateral salpingo-oophorectomy or partial oophorectomy (Colombo *et al.* (1994), Morice *et al.*, 2001, Schilder *et al.*, 2002 and Zanetta *et al.*, 1997). To prevent recurrence of disease, most doctors recommend surgery to remove the remaining ovarian tissue when a woman no longer plans to have children. Treatment for late stage ovarian low-malignant potential-tumor may be hysterectomy or bilateral salpingo-oophorectomy. A lymph node dissection may also be performed. For recurrent ovarian cancer, surgery is always necessary before chemotherapy starts (Fader & Rose, 2007).

Chemotherapy

Chemotherapy is a treatment that uses drugs to stop the growth of cancer cells, either by killing the cells or by stopping them from dividing. IV chemotherapy, meaning drugs

entering the bloodstream by mouth or by injection, is the standard procedure for OC patients. Chemotherapy can cause unpleasant side effects, but any that occur can often be well controlled with medicines (Covens *et al.*, 2002).

The length of chemotherapy varies between 9 – 24 weeks. Chemotherapy is given to OC patients in cycles. Each cycle normally is 3 – 4 weeks. How many cycles of chemotherapy is optimal was compared in some studies. Bell *et al.* (2006) compared the efficacy of 3 cycles and 6 cycles of treatment. Of the 427 eligible early stage OC patients sampled, 213 patients were given 3 cycles of chemotherapy (study group) and 214 patients were given 6 cycles of chemotherapy (control group). The authors observed that the recurrence rate for 6 cycles was 24% lower (95% confidence interval: 0.51 – 1.13), and the overall mortality rate was similar for two groups (95% confidence interval: 0.662 – 1.57). This paper concluded that 6 cycles of chemotherapy do not have a significantly better recurrence rate than 3 cycles of chemotherapy, and the former is associated with more toxicity. For late stage OC, 8 cycles of chemotherapy might be applied occasionally (Dizon *et al.*, 2006).

Radiotherapy

Some studies, e.g. Thomas (1994) and Vergote & Trimbos (2003), found radiotherapy to be an effective treatment. However, it is rarely used in treating OC patients because of its strong side effects, especially bowel obstruction. Radiotherapy is normally given to recurrent OC patients when other treatment options are no longer appropriate. It is also known as palliative radiotherapy (Thomas, 1994).

Hormonal Therapy

Hormonal therapy is still in testing currently. Zhang *et al.* (2007) introduced this new therapy for OC. Hormonal therapy is commonly called Provera®. It is a man-made drug

that is similar to the female hormone progesterone. They act as chemical messengers and help to control the activity of cells and organs.

6.3.9 Standard Treatment for Ovarian Cancer Patients

Overall the treatment options for OC are quite standardised, i.e. surgery and/or chemotherapy according to the brief introduction in Section 6.3.8. The standard treatments can be summarised as follows:

(a) Newly-diagnosed OC

For OC diagnosed at stage IA and IB, chemotherapy would not provide further benefits after the surgery (Högberg, Carstensen & Simonsen, 1993 and FIGO, 2000). Therefore, surgery not followed by chemotherapy is recommended.

For OC diagnosed at late stages, surgery followed by chemotherapy is normally recommended. Platinum-based drugs, e.g. carboplatin, together with paclitaxel, is normally recommended.

(b) Recurrent OC

For recurrent OC, surgery followed by chemotherapy is recommended. Different chemotherapy agents from first-line treatment are normally used. Palliative treatment might follow after these treatments are no longer working.

6.3.10 A Life History Model of a Woman at Risk of OC

Sections 6.3.1 to 6.3.9 introduced some important features which we need to incorporate into our model. We extended the CMIB model in Figure 6.1, and present the new model in Figure 6.2. We adapt this model from Lu, Macdonald & Waters (2008), which similarly modelled the life history of a woman at risk of BC. However, we should notice the

significant difference between OC and BC, e.g. BCs are more easily spotted at earlier stage, whereas OCs are more likely to be diagnosed at late stage.

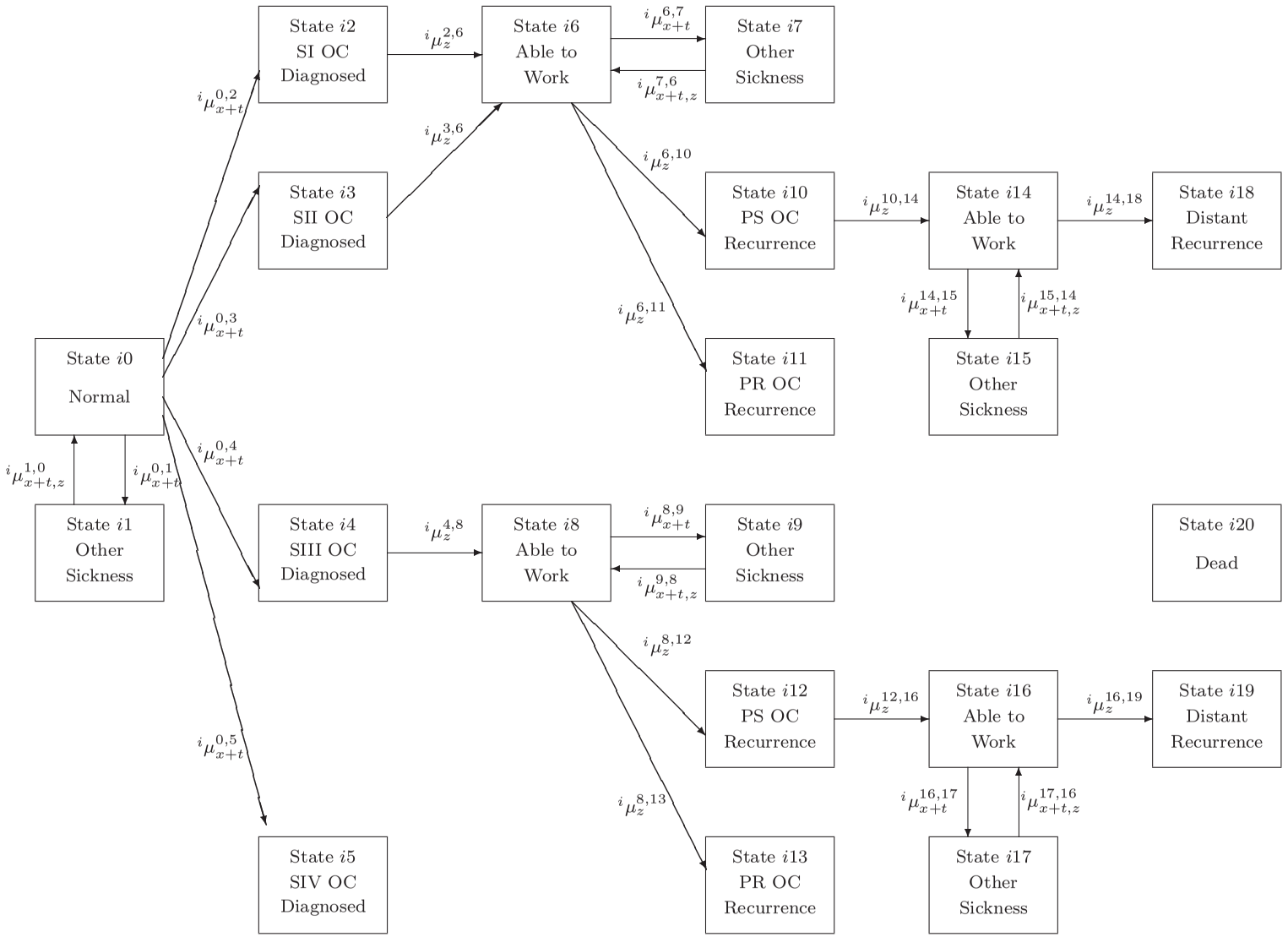

Figure 6.2: A semi-Markov model of the life history of a woman at risk of ovarian cancer with genotype i. PS OC Recurrence stands for platinum-sensitive recurrent ovarian cancer. PR OC Recurrence stands for platinum-resistant recurrent ovarian cancer.

We summarise some important features of the model in Figure 6.2 as follows:

(a) This model is semi-Markov, because some intensities, e.g. ${}^{i}\mu^{1,0}_{x+t,z}$ and ${}^{i}\mu^{7,6}_{x+t,z}$, depend on both age and duration, and some others, e.g. ${}^{i}\mu^{2,6}_{z}$ and ${}^{i}\mu^{3,6}_{z}$, depend on duration z only. All states have the access to the 'Dead' state.

(b) The genotype i indicates that the women is either a non-mutation carrier or a BRCA1/2 mutation carrier.

(c) New ovarian cancer patients can be diagnosed at different stages. Since the treatment for OC at stage IA and IB is different from that for OC at other stages, we therefore divide the diagnosis of OC into four different states as shown in Figure 6.2, where SI stands for stage IA and IB, SII for stage IC and II, SIII for stage III, and SIV for stage IV. See Section 6.3.9 for details.

(d) In the model, PS OC Recurrence stands for the platinum-sensitive recurrent ovarian cancer, and PR OC Recurrence stands for the platinum-resistant recurrent ovarian cancer (See Section 6.3.2).

(e) A healthy policyholder, without any previous medical history of OC, might be diagnosed at different stages, known as the initial diagnosis of OC. After recovering from the initial diagnosis of OC, the policyholder is still at risk of PS OC Recurrence or PR OC Recurrence, known as the first cycle of recurrence of OC. Afterwards, second, third, ..., cycle of recurrence could follow. In this model, for simplicity and mathematical consideration, we assume all the policyholders can have at most two cycles of recurrence. In reality, the second cycle of recurrence could also be either PS OC or PR OC Recurrence. In this model, we simply assume the policyholder, contracting the second recurrence, falls into a terminal state, the 'Distant Recurrence' state, which does not differentiate PS OC or PR OC Recurrence.

6.4 The Tumor Detection Model

In Section 6.3, we introduced some features of OC and presented a life history model for women at risk of OC. From this section onwards, we will parameterise the model in Figure 6.2. We assume that ovarian cancer can be detected at four different states, namely SI, SII, SIII and SIV (see Section 6.3.10 (c)). Adopting the method from Lu, Macdonald & Waters (2008), we used the SEER Program to calculate the tumour detection rates ${}^{i}\mu^{0,2}_{x+t}$, ${}^{i}\mu^{0,3}_{x+t}$, ${}^{i}\mu^{0,4}_{x+t}$ and ${}^{i}\mu^{0,5}_{x+t}$.

(a) We used the SEER data registered between 1973 and 2002. We only consider the first matching record for each person, so that recurrent OC cases are removed. Table 6.1 shows the rate of detection for OC at each stage, in which: (a) the 'Unstaged' column refers to records without this information or that are hard to classify; and (b) the 'rates' are based on the central exposure method. Here, to make use of the majority of the data, we assume that the 'Unstaged' column is distributed across the other four column *pro rata*. Table 6.2 shows the modified detection rates of OC at each stage for the general population. Figure 6.3 shows the modified detection rates of OC at each stage. These crude rates are fitted to truncated Gamma functions displayed as follows:

$$ {}^{POP}\mu^{0,2}_{x+t} = \frac{0.005818}{\Gamma(4.681)} \cdot 0.05087^{4.681} \cdot \exp(-0.05087x) \cdot x^{3.681} \tag{6.1} $$

$$ {}^{POP}\mu^{0,3}_{x+t} = \frac{7.713 \times 10^{-10}}{\Gamma(6.237395)} \cdot 0.05334^{6.237} \cdot \exp(-0.05334x) \cdot x^{5.237} \tag{6.2} $$

$$ {}^{POP}\mu^{0,4}_{x+t} = \frac{0.006533}{\Gamma(12.96)} \cdot 0.1509^{12.96} \cdot \exp(-0.1509x) \cdot x^{11.96} \tag{6.3} $$

$$^{POP}\mu^{0,5}_{x+t} = \frac{0.01010}{\Gamma(12.49)} \cdot 0.1272^{12.49} \cdot \exp(-0.1272x) \cdot x^{11.49}. \tag{6.4}$$

(b) For BRCA1/2 mutation carriers, the detection rates of specific types of OC were not reported in the medical literatures. However, the overall onset rates of OC for mutation carriers were reported by Antoniou *et al.* (2001) and fitted to truncated Gamma functions in Gui *et al.* (2006). Therefore, taking these onset rates to represent the detection of all types of OC, we assume overall onset rates are composed of the same proportion of each type of OC as their age-matched counterparts in the general population. For example, for BRCA1 mutation carriers, the detection rate of SI OC is:

$$^{BRCA1}\mu^{0,2}_{x+t} = \mu^{BC,BRCA1}_{x+t} \times \frac{^{POP}\mu^{0,2}_{x+t}}{^{POP}\mu^{0,2}_{x+t} + {}^{POP}\mu^{0,3}_{x+t} + {}^{POP}\mu^{0,4}_{x+t} + {}^{POP}\mu^{0,5}_{x+t}}. \tag{6.5}$$

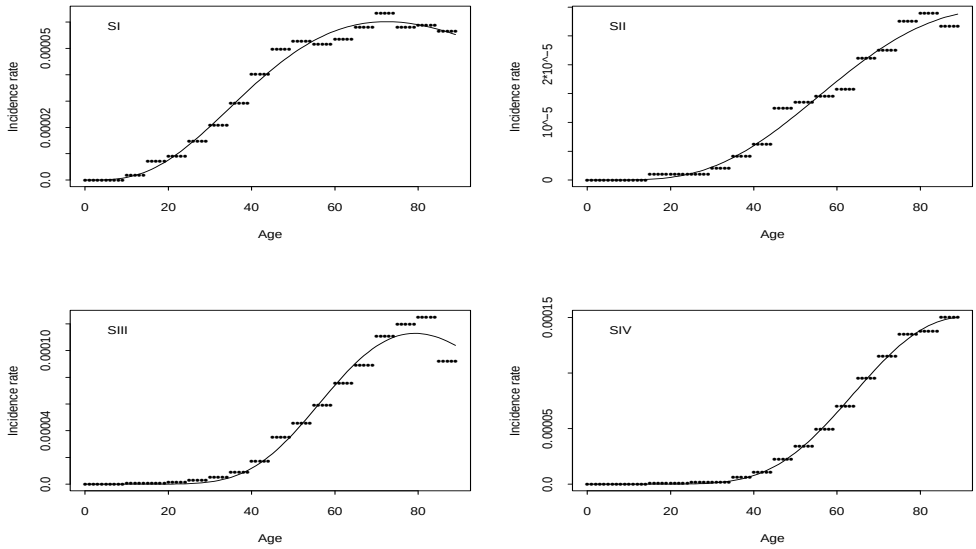

Figure 6.3: Crude estimates and fitted truncated Gamma functions of the detection rates for OC at each stage based on Table 6.2. Source: SEER Reg Public-use database (1973 – 2002)

Table 6.1: The detection rates of OC at each stage, based on the SEER Reg Public-use database. COUNT is the number of OC cases detected during the age interval and TE is the total exposure to risk. Rates are per 100,000 women years.

	SI OC		SII OC		SIII OC		SIV OC		Unstaged		
Age	COUNT	Rate	COUNT	Rate	COUNT	Rate	COUNT	Rate	COUNT	Rate	TE
00	1	0	0	0	0	0	0	0	0	0	5706482
01–04	2	0	1	0	0	0	0	0	0	0	22234214
05–09	13	0	2	0	5	0	1	0	0	0	27876958
10–14	56	0.2	9	0	17	0.1	10	0	7	0	28664877
15–19	219	0.7	16	0.1	41	0.1	15	0.1	8	0	29088143
20–24	284	0.9	32	0.1	56	0.2	24	0.1	10	0	29983883
25–29	449	1.4	45	0.1	85	0.3	49	0.2	35	0.1	31630904
30–34	647	2	73	0.2	155	0.5	65	0.2	31	0.1	31863438
35–39	856	2.8	113	0.4	277	0.9	175	0.6	63	0.2	29720429
40–44	1057	3.9	160	0.6	468	1.7	298	1.1	59	0.2	27360320
45–49	1176	4.8	281	1.2	820	3.4	539	2.2	90	0.4	24273203
50–54	1125	5.1	281	1.3	948	4.4	709	3.3	111	0.5	21705304
55–59	941	5	265	1.4	1067	5.7	913	4.8	105	0.6	18883553
60–64	823	5.1	253	1.5	1175	7.2	1098	6.7	163	1	16381303
65–69	792	5.5	298	2	1225	8.4	1311	9	213	1.5	14588298
70–74	740	5.9	264	2.1	1296	10.3	1337	10.7	277	2.2	12534262
75–79	545	5.3	253	2.5	1124	10.9	1263	12.3	318	3.1	10268158
80–84	367	5.1	183	2.5	787	10.8	868	11.9	348	4.8	7315136
85+	266	4	124	1.9	435	6.5	707	10.6	643	9.6	6673631

Table 6.2: The modified detection rates of OC at each stage, with unstaged cased being redistributed pro-rata, based on the SEER Reg Public-use database. COUNT is the number of OC cases detected during the age interval and TE is the total exposure to risk. Rates are per 100,000 women years.

	SI OC		SII OC		SIII OC		SIV OC		
Age	COUNT	Rate	COUNT	Rate	COUNT	Rate	COUNT	Rate	TE
00	1	0	0	0	0	0	0	0	5706482
01–04	2	0	1	0	0	0	0	0	22234214
05–09	13	0	2	0	5	0	1	0	27876958
10–14	60	0.2	10	0	18	0.1	11	0	28664877
15–19	225	0.7	16	0.1	42	0.1	15	0.1	29088143
20–24	291	0.9	33	0.1	57	0.2	25	0.1	29983883
25–29	474	1.5	48	0.1	90	0.3	52	0.2	31630904
30–34	668	2.1	75	0.2	160	0.5	67	0.2	31863438
35–39	894	2.9	118	0.4	289	0.9	183	0.6	29720429
40–44	1088	4	165	0.6	482	1.7	307	1.1	27360320
45–49	1214	5	290	1.2	846	3.5	556	2.3	24273203
50–54	1166	5.3	291	1.3	982	4.6	735	3.4	21705304
55–59	972	5.2	274	1.4	1102	5.9	943	5	18883553
60–64	863	5.3	265	1.6	1232	7.6	1151	7	16381303
65–69	839	5.8	316	2.1	1297	8.9	1388	9.5	14588298
70–74	796	6.3	284	2.3	1395	11.1	1439	11.5	12534262
75–79	599	5.8	278	2.7	1236	12	1389	13.5	10268158
80–84	425	5.9	212	2.9	911	12.5	1005	13.8	7315136
85+	378	5.7	176	2.7	618	9.2	1004	15	6673631

6.5 Rates of Recovery Based on Treatment Options

Usually, patients will receive treatment immediately after the diagnosis of OC and, after primary treatment followed by some rest period, they may return to work. However, OC can recur in several forms and hence there is no real 'recovery'. Here, to be consistent with the current IPI literature, 'recovery' is claim termination. It means that the woman's physical condition improves to allow her to work and she stops receiving IPI benefit, but she is still at risk of OC recurrence.

In medical studies, aiming to test the efficacy of different chemotherapies, as we will see in the subsequent sections, 'recovered' patients were randomly assigned to different testing groups, so that their cancer experience and survival experience could be compared statistically. In these medical studies, 'recovery' stands for patients having their tumours removed by surgery, but not completing the subsequent chemotherapy. However, we will define 'recovery' as being cancer-free and having completed treatment, including surgery and chemotherapy. Our definition is consistent with the definition in the context of IPI.

In the model in Figure 6.2, the recovery rate rates are functions of the sickness duration z only, because treatment procedures are highly standardized as introduced in Section 6.3.8 and 6.3.9. Therefore, we show the recovery rate from OC as follows, based on treatment options:

(a) For OC patients diagnosed at stage IA and IB, who only require surgery not followed by chemotherapy, the average period of disability is $2-4$ weeks. We assume that patients are normally given $1-2$ weeks rest period both before and after operation (Lu, Macdonald & Waters, 2008). We also assume that patients return to work evenly between the 2nd week and 4th week after the treatment, which implies a uniform distribution of the sickness duration z between 2nd week and 4th week. Hence the recovery rate ${}^{i}\mu_z^{2,6}$ could be expressed as:

$$\mu_z^{Recovery,\ Surgery} = \begin{cases} \frac{13}{1-13z} & 2/52 \leq z \leq 4/52 \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

(b) For OC patients who require surgery followed by chemotherapy, including both newly-diagnosed cases and recurrent cases, the average period of disability is 11 – 28 weeks. We assume that patients are normally given 1 – 2 weeks rest period both before and after the treatment, chemotherapy starts immediately after the surgery, and the chemotherapy normally lasts for 9 – 24 weeks. The recovery rates, ${}^i\mu_z^{3,6}$, ${}^i\mu_z^{4,8}$, ${}^i\mu_z^{10,14}$ and ${}^i\mu_z^{12,16}$, are:

$$\mu_z^{Recovery,\ Chemotherapy} = \begin{cases} \frac{13}{7-13z} & 11/52 \leq z \leq 28/52 \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

6.6 Recurrence Rates

After OC patients recover, they are still at risk of recurrent OC. As shown in the model in Figure 6.2, we assume only two cycles of recurrent OC after recovering from first-line treatment. In the first cycle, the recurrent OC could either be platinum-sensitive recurrent OC (PS OC Recurrence), or platinum-resistant recurrent OC (PR OC Recurrence). The life history of an OC patient recovering from early stage OC (stage I and II) is different from that of one recovering from late stage OC (stage III and IV). The former case is studied in Section 6.6.1, and the latter in Section 6.6.2. In the second cycle, we assume all OC patients will contract distant recurrence, which does not differentiate PS from PR. The corresponding recurrence rate is studied in Section 6.6.3.

6.6.1 Recurrence Rate for Early Stage Ovarian Cancer Patients

Here 'early stage' OC stands for stage I and II OC. We review three important prospective studies: ICON (2003), Trimbos *et al.* (2003) and ICON1 & EORTC–ACTION (2003). ICON1 & EORTC–ACTION (2003) was the combination of the two earlier similar studies. ICON1 & EORTC–ACTION (2003) contained the largest number of sampled patients in the study. Therefore, we choose to use the results presented in this paper. Some important features of this paper are as follows:

(a) The total number of patients included was 925 (477 in ICON (2003) and 448 in Trimbos *et al.*, (2003)).

(b) Patients sampled were randomly assigned to receive either chemotherapy immediately after surgery (study group; total number 465) or no chemotherapy (control group; total number 460).

(c) For all patients, all tumours were removed through surgery before randomization. The study group received 6 cycles of chemotherapy at intervals of 3 weeks.

(d) The median ages of both the study group and control group were 55 years, and the median follow-up was around 50 months.

We are interested in the Kaplan-Meier estimate of the disease-free survival probability in ICON1 & EORTC–ACTION (2003). When using this estimate, we need to pay attention to the following points:

(a) For our IPI model in Figure 6.2, 'recovery' is defined as patients completing treatment after surgery. However, observation in ICON1 & EORTC–ACTION (2003) does not start from the recovery as so defined, but from randomization just after surgery. Therefore, we are only interested in the Kaplan-Meier estimates for the

period between completion of chemotherapy and the end of observation. Based on the Kaplan-Meier estimates of the disease-free survival probability in ICON1 & EORTC–ACTION (2003), we observe that the probability of being cancer-free and alive during the period of chemotherapy is 0.98696. Therefore, we need to scale up the Kaplan-Meier estimate of interest by the reciprocal of this probability. We denote the adjusted disease-free survival probability as $S_z^{Early,\ Recurrence}$. In Figure 6.4 (top), we show the adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability. We fit the Gamma function using unweighted least squares as follows:

$$S_z^{Early,\ Recurrence} = 1 - \int_0^z 0.02058^{0.6599} \cdot t^{(0.6599-1)} \cdot e^{-0.02058t} dt \tag{6.8}$$

(b) We assume the recurrence experience is the same for patients recovering from the 'SI OC' or 'SII OC' states as shown in Figure 6.2.

(c) In this study, the events of interest are recurrent OC or death from any cause. We denote the recurrence rate as $\mu_z^{Early,\ Recurrence}$, and the duration dependent mortality rate before recurrence arises as $\mu_z^{Early,\ Pop}$, then we have:

$$S_z^{Early,\ Recurrence} = \exp\left(-\int_0^z (\mu_t^{Early,\ Recurrence} + \mu_t^{Early,\ Pop}) dt\right). \tag{6.9}$$

The intensity $\mu_z^{Early,\ Recurrence}$ can be solved once we know $\mu_z^{Early,\ Pop}$.

(d) The intensity $\mu_z^{Early,\ Pop}$ is hard to estimate without accessing the original data. Here we assume the mortality experience of the sampled patients before recurrence occurs is close to that of a normal population, because excessive mortality is accounted for by the later recurrence event. This assumption is consistent with the mortality assumption (see Chapter 7). We do not know the exact age and length of

follow-up of each patient, but the study disclosed that the median age of the study group is 55 and median follow-up is around 50 months (approximately 4.17 years). The age-dependent population mortality is that of ELT15, denoted as $\mu_x^{Standard}$. Therefore we assume

$$\mu_z^{Early,\ Pop} = \frac{\int_0^{4.17} \mu_{55+t}^{Standard} dt}{4.17}. \tag{6.10}$$

It means that $\mu_z^{Early,\ Pop}$ is treated as an average mortality rate for the sampled patients during the follow-up before recurrence occurs. Here $\mu_z^{Early,\ Pop}$ as calculated as 0.00542374. Figure 6.4 (bottom) shows the recurrence rate after the adjustment for mortality. Hence we can estimate $\mu_z^{Early,\ Recurrence}$, where z stands for the duration of being in the 'Able To Work' state.

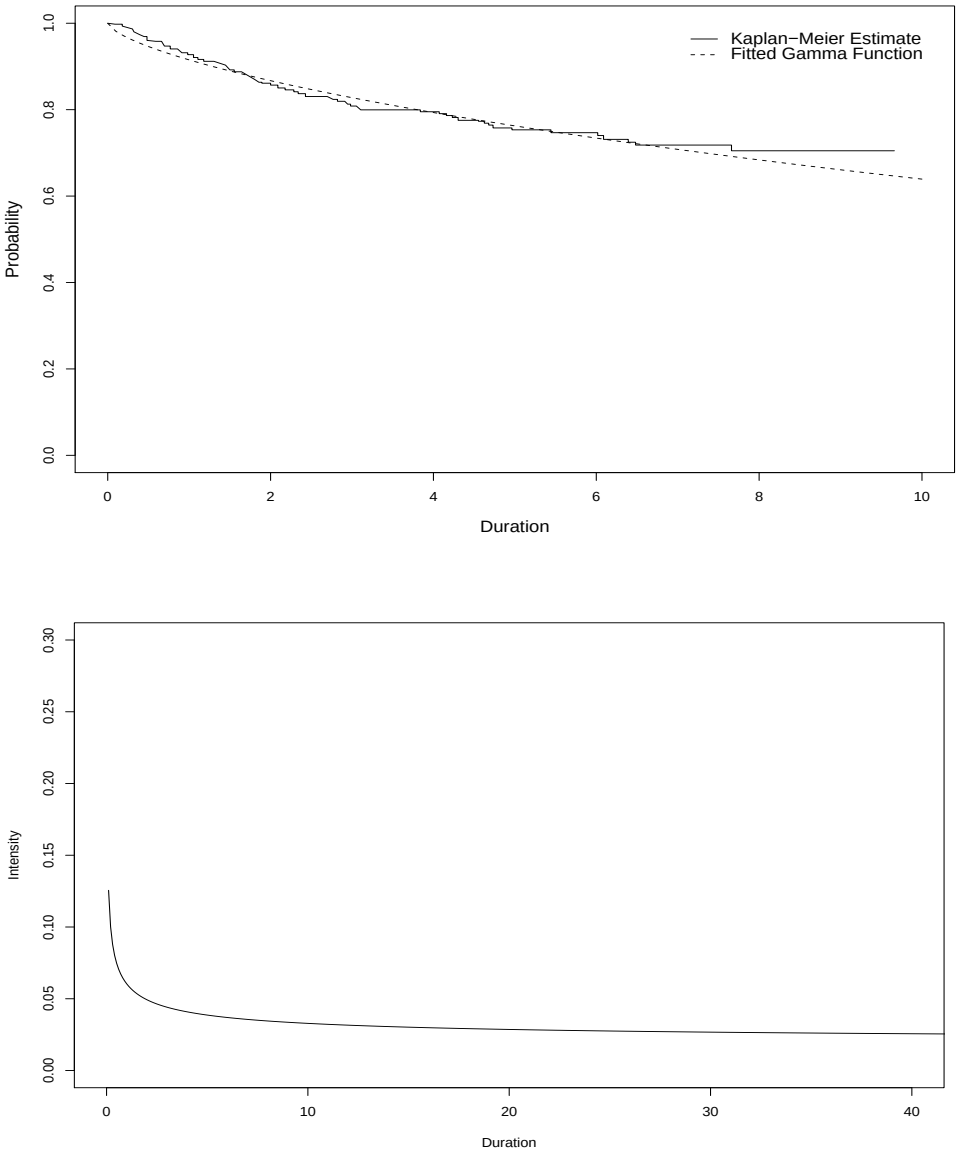

Figure 6.4: Adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability for patients recovered from early OC cancers (top), and the derived recurrence rate before adjusting for mortality (bottom). Data source: ICON1 & EORTC–ACTION (2003)

When recurrence occurs, a person, who recovered from early stage OC, will either enter the 'PS OC Recurrence' or the 'PR OC Recurrence' state. We have defined in Section 6.3.2 that the threshold between PS OC Recurrence and PR OC Recurrence is 6 months (0.5 year). Therefore we have:

$$ {}^{i}\mu_z^{6,10} = \mu_z^{Early,\ Recurrence} \times \mathbb{I}_{z \geq 0.5}, \tag{6.11} $$

$$ {}^{i}\mu_z^{6,11} = \mu_z^{Early,\ Recurrence} \times \mathbb{I}_{z < 0.5}, \tag{6.12} $$

where $\mathbb{I}$ is such an indicator function for the threshold.

6.6.2 Recurrence Rate for Late Stage Ovarian Cancer Patients

Stage IV OC patients are assumed to remain in the 'OCSIV' state in Figure 6.2 until death. Therefore, we do not consider the recurrence rate for patients recovering from stage IV OC. In order to estimate the recurrence rate after recovering from stage III OC, denoted as $\mu_z^{Recurrence,\ Late}$, we located Dizon *et al.* (2006). This study sampled 122 stage III OC patients who had tumours removed by surgery. These patients were randomly assigned to receive 6 cycles ($n = 64$) or 8 cycles ($n = 38$) of carboplatin and paclitaxel as adjuvant treatment. The disease-free survival probabilities were calculated for both 6 cycles and 8 cycles groups using Kaplan-Meier methods. The observation starts from randomization. We choose to use the Kaplan-Meier estimate of disease-free survival probability for patients receiving 6 cycles of chemotherapy. As in Section 6.6.1, we should pay attention to the following:

(a) Observation in Dizon *et al.* (2006) does not start from the recovery as so defined, but from randomization just after surgery. Based on the Kaplan-Meier estimates of the disease-free survival probability, we observe that the probability of being cancer-free and alive during the period of chemotherapy is 0.97645. Therefore, we need to

adjust the observed Kaplan-Meier estimate by this probability. See Section 6.6.1 for details. We denote the adjusted disease-free survival probability as $S_z^{Late,\ Recurrence}$. In Figure 6.5 (top), we show the adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability. We fit the Kaplan-Meier estimate with a Gamma function using unweighted least squares as follows:

$$S_z^{Late,\ Recurrence} = 1 - \int_0^z 0.4415^{0.7989} \cdot t^(0.7989 - 1) \cdot e^{-0.4415 \cdot t} dt \tag{6.13}$$

(b) In this study, the events of interest are recurrent OC or death from any cause. If we denote the recurrence rate as $\mu_z^{Late,\ Recurrence}$, and the duration dependent mortality rate before recurrence occurs as $\mu_z^{Late,\ Pop}$, then we have:

$$S_z^{Late,\ Recurrence} = \exp\left(- \int_0^z (\mu_t^{Late,\ Recurrence} + \mu_t^{Late,\ Pop}) dt \right) \tag{6.14}$$

The intensity $\mu_z^{Late,\ Recurrence}$ can be solved for once we know $\mu_z^{Late,\ Pop}$.

(c) The intensity $\mu_z^{Late,\ Pop}$ is hard to estimate without accessing the original data. We use the same assumption about mortality as in Section 6.6.1. The study disclosed that the median age of the study group is 60.2 and median follow-up is around 13 months (approximately 1.08 years). The age-dependent population mortality is that of ELT15, denoted as $\mu_x^{Standard}$. Therefore we assume

$$\mu_z^{Late,\ Pop} = \frac{\int_0^{1.08} \mu_{60.2+t}^{Standard} dt}{1.08}. \tag{6.15}$$

It means that $\mu_z^{Late,\ Pop}$ is treated as an average mortality rate for the sampled patients during the follow-up before occurrence occurs. We calculate that $\mu_z^{Late,\ Pop} = 0.00833$. Hence we can estimate $\mu_z^{Late,\ Recurrence}$, where z stands for the duration of

being in the ‘Able To Work’ state. Figure 6.5 (bottom) shows the recurrence rate after adjustment for mortality.

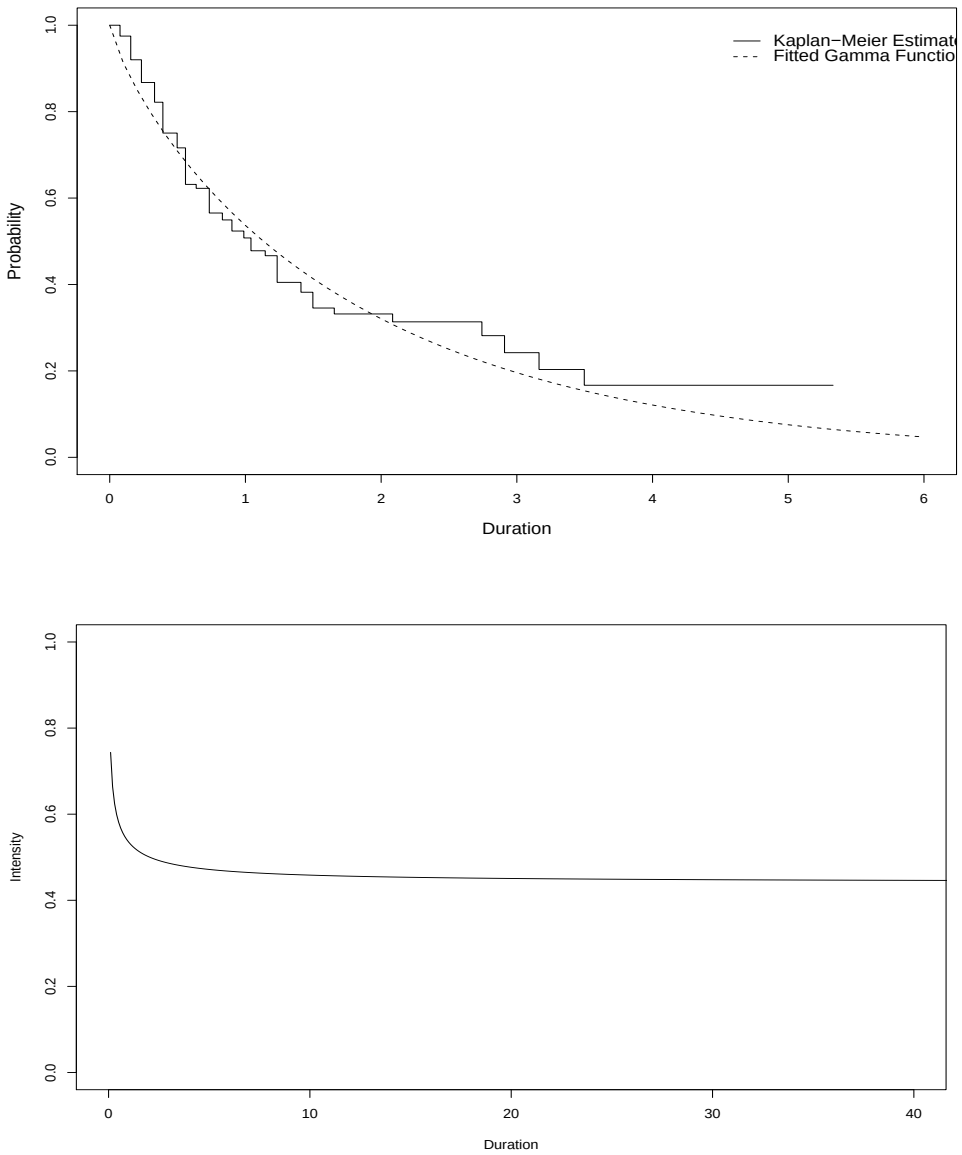

Figure 6.5: Adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability for patients recovered from late OC (top), and the derived recurrence rate adjusted for mortality (bottom). Data source: Dizon *et al.* (2006)

.

When recurrence occurs, a person, who recovered from late stage OC, will either enter into the 'PS OC Recurrence' or the 'PR OC Recurrence' state. We have defined in Section 6.3.2 that the threshold between PS OC Recurrence and PR OC Recurrence is 6 months (0.5 year). Therefore we have:

$$ {}^{i}\mu_z^{8,12} = \mu_z^{Late,\ Recurrence} \times \mathbb{I}_{z \geq 0.5}, \tag{6.16} $$

$$ {}^{i}\mu_z^{8,13} = \mu_z^{Late,\ Recurrence} \times \mathbb{I}_{z < 0.5}, \tag{6.17} $$

where $\mathbb{I}$ is an indicator function for the threshold.

6.6.3 Distance Recurrence Rate for Relapsed Ovarian Cancer Patients

Patients who have recovered from early stage OC might suffer PS OC Recurrence or PR OC Recurrence. In the case of the PR OC Recurrence, we assume patients will remain in the 'PR OC Recurrence' state until death, because patients become more platinum-resistant to the subsequent treatments. In the case of the PS OC Recurrence, we assume patients are still at risk of recurrent OC (called 'distant recurrence' in Figure 6.2) after recovery. Therefore, we only consider the recurrence rate after recovering from PS OC Recurrence, i.e. ${}^{i}\mu_z^{14,18}$. Ideally, we should sample the patients, who have firstly recovered from early stage OC and then PS OC Recurrence, in order to estimate intensity ${}^{i}\mu_z^{14,18}$. In Section 6.3, we mentioned that OC patients are less likely to be detected at an early stage, hence it is also hard to sample eligible OC patients, who satisfy the above criteria. Actually, we can not locate any paper on this topic. Therefore, we decide to use the recurrence rate after recovering from stage III, i.e. $\mu_z^{Recurrence,Late}$ to represent intensity ${}^{i}\mu_z^{14,18}$. In Section 6.6.2, we have estimated the intensity $\mu_z^{Recurrence,Late}$. Therefore we assume that ${}^{i}\mu_z^{14,18} = \mu_z^{Recurrence,\ Late}$.

Patients who have recovered from late stage OC might also suffer PS OC Recurrence or PR OC Recurrence. We also assume that patients in the 'PR OC Recurrence' state will remain in this state until death. Hence only the recurrence rate $\mu_z^{Recurrence,\ Distant} = {}^{i}\mu_z^{16,19}$ needs to be estimated. We chose to use the result reported in Pfisterer *et al.* (2006). In this intergroup trial, 356 platinum-sensitive recurrent patients were randomly assigned to receive either gemcitabine plus carboplatin (group A, $n = 178$) or carboplatin alone (group B, $n = 178$). Six cycles of chemotherapy were given to the patients at intervals of 21 days. 80% of all patients had an initial diagnosis of late stage OC. Disease-free survival curves were calculated using Kaplan-Meier methods. The median follow-up was 8.5 months and 6 months for group A and B, respectively. As in Section 6.6.1 and 6.6.2, we should pay attention to the following:

(a) Observation in Pfisterer *et al.* (2006) does not start from the recovery as so defined, but from randomization just after surgery. Based on the Kaplan-Meier estimates of the disease-free survival probability, we observe that the probability of being cancer-free and alive during the period of chemotherapy is 0.95413. Therefore, we need to adjust the observed Kaplan-Meier estimate by this probability. See Section 6.6.1 for details. We denote the adjusted disease-free survival probability as $S_z^{Distant,\ Recurrence}$. In Figure 6.6 (top), we show the adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability. We fit the Kaplan-Meier estimate with a Gamma function using unweighted least squares as follows:

$$S_z^{Distant,\ Recurrence} = 1 - \int_0^z 3.422^{2.710} \cdot t^{(2.710-1)} \cdot e^{-3.422 \cdot t} dt \tag{6.18}$$

(b) In this study, the events of interest are recurrent OC or death from any causes. If we denote the recurrence rate as $\mu_z^{Distant,\ Recurrence}$, and the duration dependent mortality rate before recurrence occurs as $\mu_z^{Distant,\ Pop}$, then we have:

$$S_z^{Distant,\ Recurrence} = \exp\left(-\int_0^z (\mu_t^{Distant,\ Recurrence} + \mu_t^{Distant,\ Pop})dt\right) \tag{6.19}$$

The intensity $\mu_z^{Distant,\ Recurrence}$ can be solved for once we know $\mu_z^{Distant,\ Pop}$.

(c) The intensity $\mu_z^{Distant,\ Pop}$ is hard to estimate without accessing the original data. We use the same assumption about mortality as in Section 6.6.1 and 6.6.2. The study disclosed that the median age of the study group is 59 and median follow-up is 8.6 months (approximately 0.717 year). The age-dependent population mortality is that of ELT15, denoted as $\mu_x^{Standard}$. Therefore we assume

$$\mu_z^{Distant,\ Pop} = \frac{\int_0^{0.717} \mu_{59+t}^{Standard} dt}{0.717}. \tag{6.20}$$

It means that $\mu_z^{Distant,\ Pop}$ is treated as an average mortality rate for the sampled patients during the follow-up before occurrence occurs. We calculate that $\mu_z^{Distant,\ Pop} = 0.00743$. Hence we can estimate $\mu_z^{Distant,\ Recurrence}$, where z stands for the duration of being in the 'Able To Work' state. Figure 6.6 (bottom) shows the recurrence rate after adjustment for mortality.

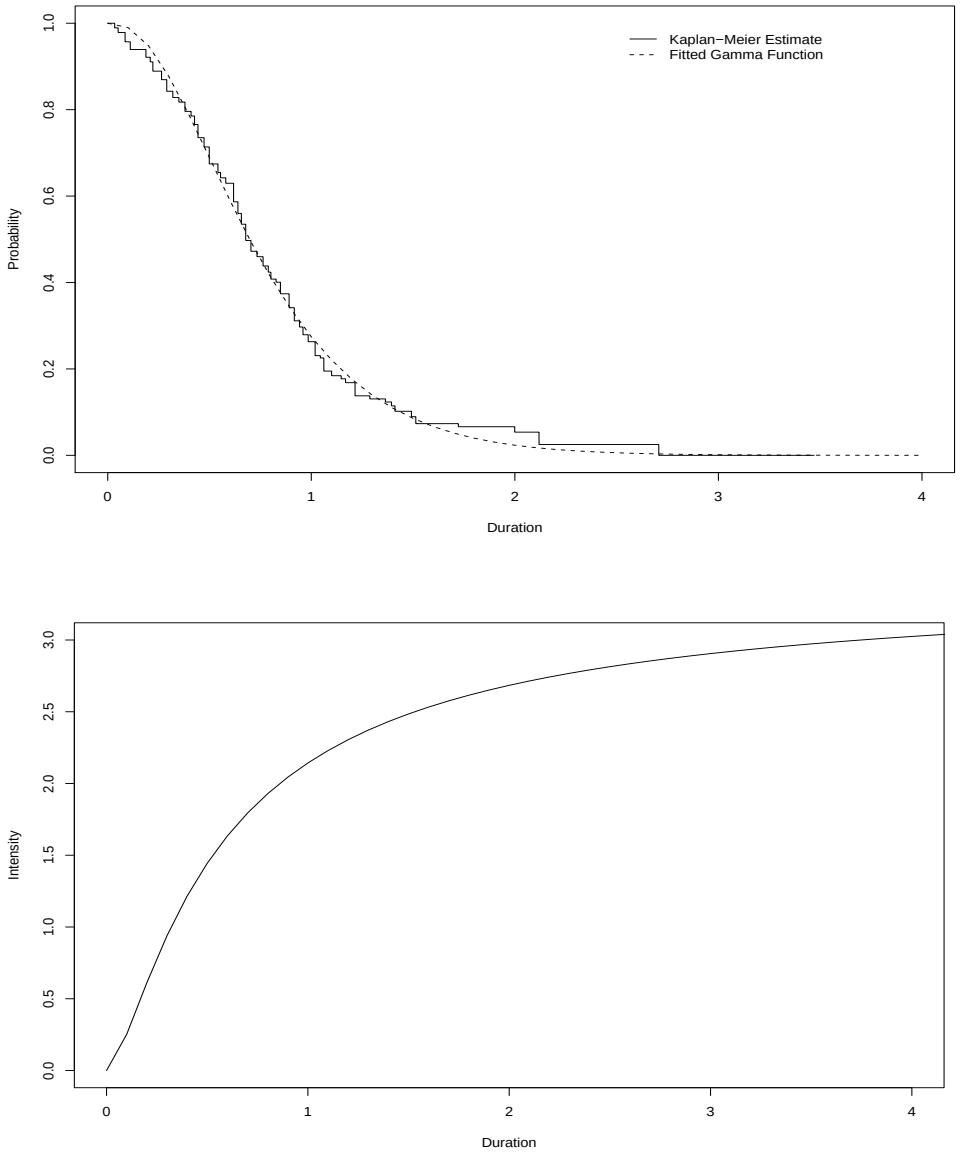

Figure 6.6: Adjusted Kaplan-Meier estimate and fitted Gamma function for the disease-free survival probability for patients recovered firstly from late stage OC and secondly from PS OC Recurrence (top), and the derived recurrence rate adjusted for mortality (bottom). Data source: Pfisterer *et al.* (2006).

Chapter 7

Application of the Life History Model to Income Protection Insurance

In Chapter 6, we constructed a life history model of a woman at risk of OC, incorporating some important features of OC, e.g. diagnosis, progression, treatment, recurrence, etc. In this chapter, we use this model to calculate the premium rates based on genetic test results or on a family history of OC. We then expand the model to an IPI market by including insurance purchasing behaviours and calculate the potential cost of adverse selection under various moratoria on the use of genetic information.

7.1 Other Intensities

7.1.1 Intensities for Sickness Other Than Ovarian Cancer

The Onset Rate of Other Sickness

CMIR 12 (1991) showed the graduated sickness inception intensities for four deferred periods (DPs):

$$\sigma_x^{R12} = \begin{cases} \exp(-1.796 + 8.803 \times 10^{-2}x \\ -2.686 \times 10^{-3}x^2 + 2.498 \times 10^{-5}x^3) & \text{for DP 1 week} \\ \exp(-4.256 + 2.392 \times 10^{-1}x \\ -6.498 \times 10^{-3}x^2 + 5.476 \times 10^{-5}x^3) & \text{for DP 4 weeks} \\ \exp(-2.722 + 1.290 \times 10^{-1}x \\ -4.240 \times 10^{-3}x^2 + 3.888 \times 10^{-5}x^3) & \text{for DP 13 weeks} \\ \exp(-0.482 - 8.434 \times 10^{-1}x \\ 9.749 \times 10^{-4}x^2) & \text{for DP 26 weeks} \end{cases} \tag{7.1}$$

The intensity σ_x^{R12} is the rate of onset at age x of all sicknesses which could result in an IPI claim (CMIR 12, 1991). These estimates were based on data collected between 1975 and 1978. In this thesis we intend to use more recent experience. CMIR 20 (2001) studied the IPI data between 1995 and 1998 and reported the modifying factors of claim inception rates, $M^{inception}$, as in Table 7.1. The factor $M^{inception}$ is the ratio of the actual IPI claim inceptions between 1995 and 1998 to those expected using the sickness inception intensities parameterised using the data between 1975 and 1978 in CMIR 12.

We should notice the distinction between sickness inception and claim inception. Firstly the former happen before the latter and secondly the former might not trigger the latter. Therefore what we are looking for is a different modifying factor, denoted as M^{sick}, which is the ratio of sickness inception intensity based on CMIR 20 (2001) to that

Table 7.1: The modifying factor $M^{inception}$ of claim inception rates based on the standard female experience for 1995–1998. Source: CMIR 20 (2001).

DP 1 week	DP 4 weeks	DP 13 weeks	DP 26 weeks
1.260	1.031	1.614	3.420

based on CMIR 12 (1991), so that the sickness inception rate based on CMIR 20 (2001), denoted as σ_x^{R20}, can be represented explicitly as:

$$\sigma_x^{R20} = M^{sick} \times \sigma_x^{R12}. \tag{7.2}$$

Policyholders making a valid IPI claim should experience the following stages:

(1) Contract sickness;

(2) Remain sick over the DP specified in the IPI policy;

(3) Make a claim.

Lu *et al.* (2008) estimated the modifying factor M^{sick} by considering the above stages. We summarise the procedure as follows:

(a) Let d be the DP in years, and $w = 52$ be the number of weeks in a year. Let $\tilde{\sigma}_x^{R12}$ and $\tilde{\sigma}_x^{R20}$ be the age-dependent claim inception rates based on CMIR 12 and CMIR 20, respectively. Let $\pi_{y,d}^{R12}$ be the probability that an individual who falls sick at age y, will remain sick for at least the deferred period d of the policy, based on CMIR 12, and let $\pi_{y,d}^{R20}$ be the corresponding value based on CMIR 20. Let $r_d(y+d)$ be the proportion of potential claims actually reported at age $y+d$ for policies with deferred period d. Denote quantities based on CMIR 12 and CMIR 20 as $r_d^{R12}(y+d)$

and $r_d^{R20}(y+d)$, respectively. So that the claim inception rate based on CMIR 12 can be expressed as:

$$\tilde{\sigma}_{y+d}^{R12} = \sigma y^{R12} \cdot \pi_{y,d}^{R12} \cdot r_d^{R12}(y+d), \tag{7.3}$$

and we can write out the corresponding equation for the claim inception rate based on CMIR 20.

(b) The modifying factor $M^{inception}$ is the ratio of $\tilde{\sigma}_x^{R20}$ to $\tilde{\sigma}_x^{R12}$ and the modifying factor M^{sick} is the ratio of σ_x^{R20} to σ_x^{R12}. Therefore we have:

$$\frac{\tilde{\sigma}_{y+d}^{R20}}{\tilde{\sigma}_{y+d}^{R12}} = \frac{\sigma_y^{R20} \cdot \pi_{y,d}^{R20} \cdot r_d^{R20}(y+d)}{\sigma_y^{R12} \cdot \pi_{y,d}^{R12} \cdot r_d^{R12}(y+d)}, \tag{7.4}$$

that is:

$$M^{inception} = M^{sick} \times \frac{\pi_{y,d}^{R20} \cdot r_d^{R20}(y+d)}{\pi_{y,d}^{R12} \cdot r_d^{R12}(y+d)}, \tag{7.5}$$

and hence:

$$M^{sick} = M^{inception} \times \frac{\pi_{y,d}^{R12} \cdot r_d^{R12}(y+d)}{\pi_{y,d}^{R20} \cdot r_d^{R20}(y+d)}. \tag{7.6}$$

(c) The definition of $\pi_{y,d}$ is as follows:

$$\pi_{y,d} = \exp\left(-\int_0^d (\rho_{y+s,s} + \nu_{y+s,s}\right). \tag{7.7}$$

This is a general equation which can be easily modified for the corresponding values based on CMIR 12 and CMIR 20. We should notice that $\pi_{y,d} = {}_dp_{y,0}^{\overline{SS}}$ in the standard actuarial notation system.

(d) The proportion $r_d(y)$ was introduced in CMIR 13 (1993) to make some allowance when comparing actual claim inceptions with those expected according to the model in CMIR 12. In CMIR 12, it was observed that the recovery rates for DP 4, 13 and 26 weeks were considerably lower than those for DP 1 week during the four weeks immediately following the end of the DP. This was explained by the possibility that some policyholders whose sickness lasts only a little beyond the end of the DP do not bother to claim benefit, called 'non-reported claims'. Following the notation in CMIR 13, we define the function $pSS^d(y, z_1, z_2) = {}_{z_2-z_1}p^{\overline{SS}}_{y+z_1,z_1}$ for DP d, that is the probability that a life who became sick at age y and is still sick at duration z_1 remains sick until duration z_2. Then we have:

$$r^d(y+d) = \frac{pSS^1(y, d, d+4/w)}{pSS^d(y, d, d+4/w)}, \tag{7.8}$$

and ${}_{z_2-z_1}p^{\overline{SS}}_{y+z_1,z1}$ can be calculated using Equation (7.7). We should notice that equation (7.8) implies that 'a little beyond the end of the DP' mentioned above is 4 weeks in CMIR 13.

(e) Note that women diagnosed with OC usually have sickness periods much longer than typical DPs and it is reasonable to assume they will claim at the end of the DP. Therefore, in respect of them we would assume $\pi_{y,d} = r^d(y) = 1$.

We calculate the proportion $r_d(y)$ for different DPs and show results in Figure 7.1.

In our model in Figure 6.2, we distinguish OC from other sicknesses. So the intensity of contracting 'other sickness', denoted as $\mu_x^{othersick}$, is represented by deducting the onset rate of OC from the intensity for females σ_x^{R20}.

$$\begin{aligned} \mu_x^{othersick} &= \sigma_x^{R20} - \mu_x^{OC} \\ &= M^{sick} \times \sigma_x^{R12} - \mu_x^{OC}. \end{aligned} \tag{7.9}$$

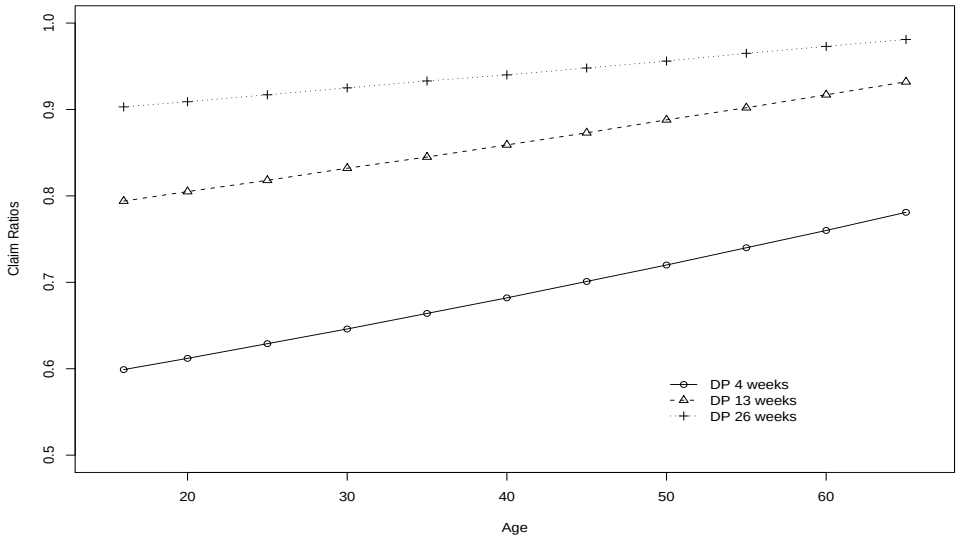

Figure 7.1: Proportion $r_d(y)$ for different DPs, based on the standard female experience for 1995–1998.

Figure 7.2 shows the resulting intensities of other sickness for different DPs, based on CMIR 20.

The Rates of Claim Recovery From Other Sicknesses

The claim recovery rates depend on the current age $x+t$ and sickness duration z. Let $\rho^{R12}_{y+z,z}$ and $\rho^{R20}_{y+z,z}$ be the claim recovery rates based on CMIR 12 and CMIR 20, respectively. CMIR 12 introduced graduations of recovery rates based on 1975–1978 data for males as follows:

$$\begin{aligned}\rho^{R12}_{y+z,z} &= r \cdot \{51.057 - 2.687(1 + 1.419 \cdot \max(4 - wz, 0)) \\ &\quad \cdot\sqrt{Z} \cdot (Y - 50)\} \cdot e^{-4.914\sqrt{Z}}, \end{aligned} \tag{7.10}$$

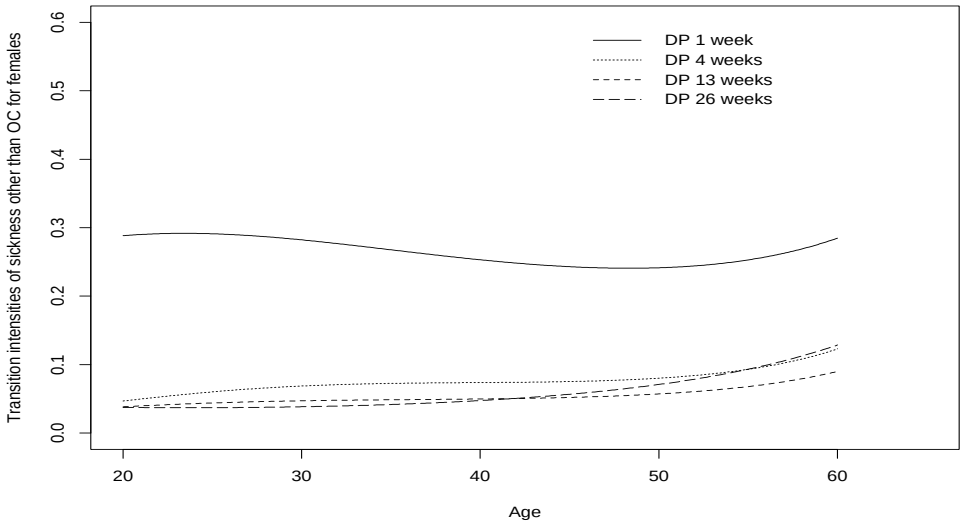

Figure 7.2: Onset rates of sickness other than OC for different DPs, based on the standard female experience for 1995–1998.

where:

$$\begin{aligned} y &= \text{the exact age at the time of falling sick;} \\ z &= \text{the sickness duration in years;} \\ wz &= \text{the sickness duration in weeks;} \end{aligned}$$

$$Y = \begin{cases} y & z \leq 5 \\ y + z - 5 & \text{otherwise} \end{cases} \tag{7.11}$$

$$Z = \begin{cases} z & z \leq 1 \\ 1 + 0.363(z-1) & 1 < z \leq 5 \\ 2.450 & \text{otherwise} \end{cases} \tag{7.12}$$

$$r = \begin{cases} \min(0.205 + 0.199(wz - DP), 1) & \text{DP} = 4,\ 13,\ 26 \text{ weeks only} \\ 1 & \text{otherwise.} \end{cases} \tag{7.13}$$

Similar to the treatment given to the sickness inception rate in Section 7.1.1, CMIR 20 introduced the modifying factors $M^{recovery}$ for females in 1995 – 1998 shown in Table 7.2, implying that the resulting recovery rates for females based on CMIR 20 are

$$\rho^{R20}_{y+z,z} = M^{recovery} \times \rho^{R12}_{y+z,z} \tag{7.14}$$

Table 7.2: The modifying factor $M^{recovery}$ of claim recovery rates based on the standard female experience for 1995–1998. Source: CMIR 20 (2001).

	DP 1	DP 4	DP 13	DP 26
$\text{DP}/w \leq z < (\text{DP} + 4)/w$	0.59	0.53	0.45	0.45
Otherwise	0.59	0.59	0.59	0.59

7.1.2 Mortality

Recall that in Section 6.3.10 all states, (except the state 'Dead') in the model in Figure 6.2 have the access to the 'Dead' state. For simplicity, we call these states the death entry states. The mortality rates from these death entry states entering into the state 'Dead' should modelled. However, these mortality rates are modelled differently depending on the natures of the death entry states. We split the death entry states into three parts as follows:

(1) Part I: The 'Normal' state or the 'Able to Work' states, i.e. State $i0$, $i6$, $i8$, $i14$ and $i16$.

Generally the persons in these death entry states are in better health status than persons in others. We model the corresponding mortality rates in Section 7.1.2.

(2) Part II: First diagnosis of OC, i.e. State $i2$ to $i5$, and Recurrence of OC, i.e. State $i10$ to $i13$ and State $i18$ to $i19$.

After contracting OC, either first diagnosis or recurrence, the persons in these death entry states are at greater risk of death depending on the severity of OC. We model the corresponding mortality rates in Section 7.1.2.

(3) Part III: The 'Other Sickness' states, i.e. States $i1$, $i7$, $i9$, $i15$ and $i17$.

The corresponding mortality rates were modelled in CMIR 12 (1991) and CMIR 20 (2001). We introduce these in Section 7.1.2.

Baseline Mortality

The basic mortality used in this study is that of English Life Tables No. 15 (ELT15), denoted as $\mu_x^{standard}$. Women in the 'Able to Work' states or 'Normal' state are just recovering from OC treatment or before any episode of OC, so these person are generally in better health status than persons in other states. Therefore the basic mortality should be applied to these persons, i.e. $\mu_x^{standard}$. However, any extra mortality caused by OC should be accounted for by later OC events, so we should remove deaths caused by OC.

In Section 3.4.2, we calculated the mortality excluding death caused by BC and OC by multiplying the basis mortality $\mu_x^{standard}$ by $(1 - r_x^{BCOC})$, where r_x^{BCOC} is the ratio of the number of deaths caused by BC and OC to the total population numbers of deaths (see Wekwete (2002) for details). We used the same method and calculated the ratio ϕ_x, which is the number of the number of deaths from OC to the total number of deaths. The ratio ϕ_x was graduated based on UK population data in 1990–1992 (O.P.C.S., 1991, 1993; O.N.S., 1999):

$$\phi_x = \begin{cases} \frac{2.028}{\Gamma(11.86)} \cdot (0.2120)^{11.86} \cdot \exp(-0.2120x) \cdot x^{10.86} & x \leq 55 \\ 0.4289 - 0.01137x + 9.897 \times 10^{-5}x^2 - 2.814 \times 10^{-7}x^3 & x > 65 \end{cases} \tag{7.15}$$

with linear interpolation between age 55 and 65. Hence the adjusted mortality is $\mu^{standard} \times (1 - \phi_x)$, shown in Figure 7.3. The adjusted mortality is applied to the intensities ${}^i\mu_x^{0,20}$, ${}^i\mu_x^{6,20}$, ${}^i\mu_x^{8,20}$, ${}^i\mu_x^{14,20}$ and ${}^i\mu_x^{16,20}$.

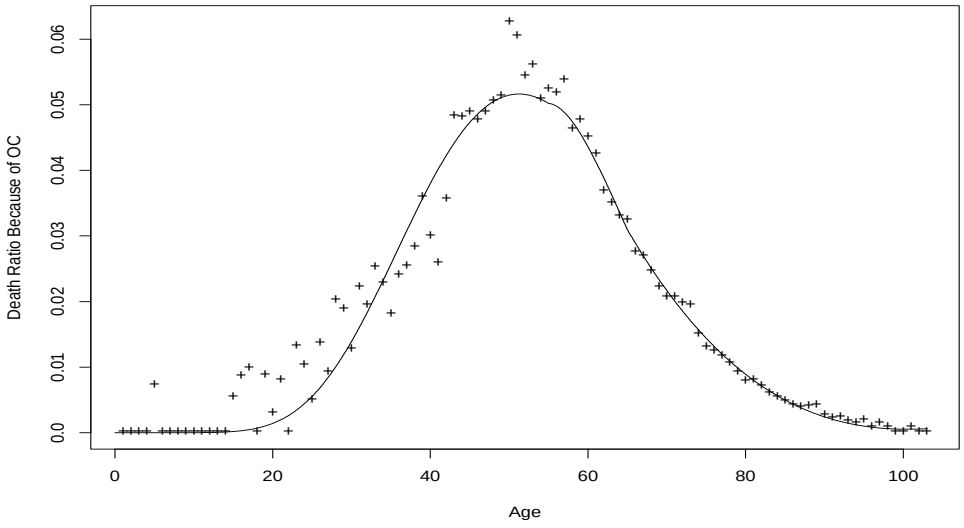

Figure 7.3: Crude and graduated proportions of total deaths that are due to OC, for females.

Mortality since Diagnosis of Ovarian Cancer

We used the SEER*Stat version 6.4.4 to calculate the mortality for women diagnosed with OC based on the SEER 9 Regs Public-use (1973–2002) database. The SEER*Stat version 6.4.4, instead of calculating the absolute mortality rates directly, has the function to calculate the relative survival rates (RSR), which is defined as follows.

Let ${}_tp_{x,z}^{cancer}$ be the t-year survival probability for an OC patient at age x with duration z since diagnosis and let $\mu_{x,z}^{cancer}$ be the corresponding force of mortality; define ${}_tp_x^{standard}$ and $\mu_t^{standard}$ to be the corresponding values for the standard population. Here we assume:

$$\mu_{x,z}^{cancer} = \mu_x^{standard} + \mu_z^{extra}, \tag{7.16}$$

where μ_z^{extra} is additional mortality for women diagnosed with OC, which we assume depends only on sickness duration z. Then the relative survival rates (RSR) for women suffering OC are defined as follows:

$$RSR = \frac{{}_tp_{x,z}^{cancer}}{{}_tp_x^{standard}} = \frac{\exp\left(-\int_0^t (\mu_{x+s}^{standard} + \mu_s^{extra})ds\right)}{\exp\left(-\int_0^t \mu_{x+s}^{standard} ds\right)} = \exp\left(-\int_0^t \mu_s^{extra} ds\right). \tag{7.17}$$

We produced the RSRs for women suffering each type of OC, and fitted them using smooth functions of the duration since diagnosis, z, as shown in Figures 7.4, 7.5 and 7.6.

Here, the standard population is the US Standard Population in 2000 and we assume similar mortality trends in respect of women at risk of OC in the US and in the UK. Therefore, the extra mortality, μ_z^{extra}, can be deduced from the fitted functions of RSRs for each stage of OC:

$$\mu_z^{extra} = \begin{cases} 0.03740 - 0.01086z + 0.002012z^2 \\ -1.865 \times 10^{-4} z^3 + 7.906 \times 10^{-6} z^4 - 1.220 \times 10^{-7} z^5 & \text{for Early Stage OC} \\ 0.1555 - 0.03756z + 0.003610z^2 \\ -1.438 \times 10^{-4} z^3 + 1.428 \times 10^{-6} z^4 + 2.375 \times 10^{-8} z^5 & \text{for Stage III OC} \\ 0.4508 - 0.09978z + 0.008329z^2 \\ -3.091 \times 10^{-4} z^3 + 4.297 \times 10^{-6} z^4 - 6.810 \times 10^{-10} z^5 & \text{for Stage IV OC.} \end{cases} \tag{7.18}$$

The early stage OC is represented by the State $i2$ and $i3$, the Stage III OC is represented by the State $i4$, and the Stage IV OC is represented by the State $i5$ in Figure 6.2.

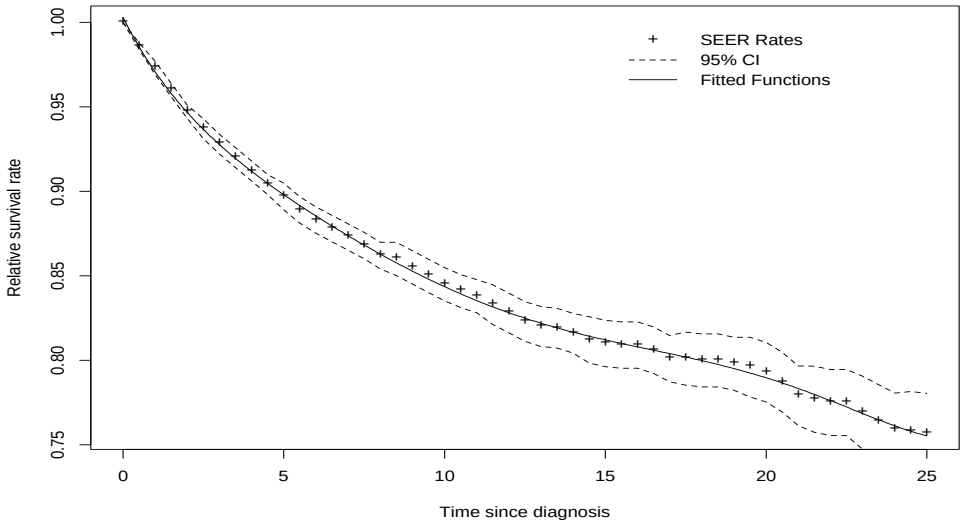

Figure 7.4: Relative survival rate for women who develop early ovarian cancer.

For women contracting recurrent OC, instead of modelling the corresponding mortality rates separately, we separated these states into three groups based on the severity of OC as follows:

(a) low-level risk group: being diagnosed at early stage of OC, i.e. State $i2$ and $i3$;

(b) medium-level risk group: being diagnosed at stage III of OC, i.e. State $i4$, or contracting PS OC Recurrence, State $i10$ and $i12$; so we used the mortality rates for women contracting Stage III OC to represent this group.

(c) advanced-level risk group: being in terminal states, i.e. States $i5$, $i11$, $i13$, $i18$ and $i19$; so we used the mortality rates for women contracting Stage IV OC to represent this group.

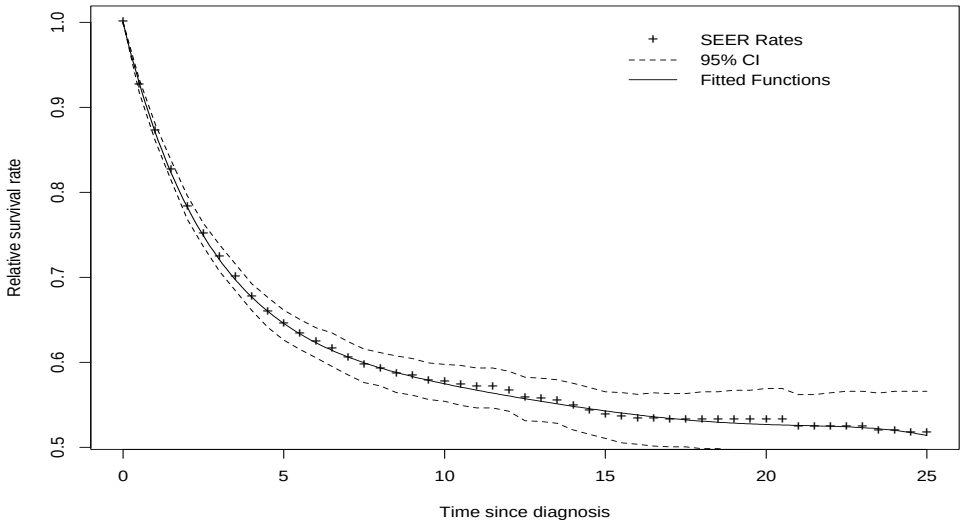

Figure 7.5: Relative survival rate for women who develop Stage III ovarian cancer

Mortality after Onset of Other Sickness

The graduated and aggregate mortality $\nu^{R12}_{y+z,z}$ was reported in CMIR 12:

$$\begin{aligned}\nu^{R12}_{y+z,z} &= \{(0.2379 - 4.819 \times 10^{-3}Y + 9.587 \times 10^{-5}Y^2) \\ &\quad \cdot \frac{\exp(-0.8747/(Z+0.3574)^{1}.6139)}{(Z+0.3574)^{2}.6139)} + 7.221 \times 10^{-3} \cdot e^{0.02435(Y+Z)}\}, \quad (7.19)\end{aligned}$$

where y, z and Y have the same definitions as in Equation 7.10,

$$Z = \begin{cases} z & z \leq 5 \\ 5 & z > 5 \end{cases} \quad (7.20)$$

Following the same treatments given to the sickness inception rate σ^{R20}_x (see Section 7.1.1) and recovery rate $\rho^{R20}_{y+z,z}$ (see Section 7.1.1), CMIR 20 reported the modifying factor

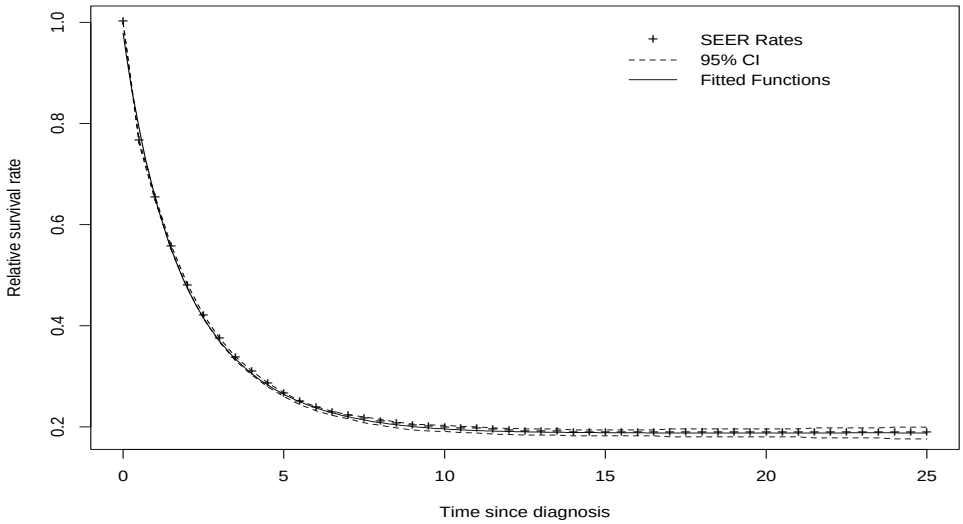

Figure 7.6: Relative survival rate for women who develop Stage IV ovarian cancer

$M^{sickmort}$, so that the corresponding mortality $\nu^{R20}_{y+z,z}$ could be derived based on $\nu^{R12}_{y+z,z}$ as follows:

$$\nu^{R20}_{y+z,z} = M^{sickmort} \times \nu^{R12}_{y+z,z}, \tag{7.21}$$

and $M^{sickmort}$ is given in Table 7.3.

Table 7.3: The modifying factor $M^{sickmort}$ of mortality rates after onset of sickness other than OC based on the standard female experience for 1995–1998. Source: CMIR 20 (2001).

DP 1 week	DP 4 weeks	DP 13 weeks	DP 26 weeks
0.15	0.34	0.33	0.26

7.1.3 The Onset Rate of a Family History Relevant to OC

In Section 3.5, we briefly introduced the onset rate of a family history relevant to BC & OC. The methodology is detailed in Gui *et al.* (2006). In Gui *et al.* (2006), the definition of family history was two first degree relatives with BC or OC before age 50, this is because mutations BRCA1 and BRCA2 carriers are at increased risks of both BC and OC. This means that family members contracting BC before age 50 indicates that other healthy family members are not only at higher risks of BC, but OC as well. In Lu *et al.* (2008), the authors defined the family history as two first degree relatives with BC only before age 50. In this chapter, as an extension of Lu, Macdonald & Waters (2008) and Lu *et al.* (2008), we ignore BC and define a family history as two first degree relatives (FDRs) with OC before age 50. This implies that we used an underestimated onset rate of family history in this thesis. A separate research work should be carried out to study the effect of genetic information relevant to both BC and OC in a combined model, in which a family history relevant to BC and OC should be used. As in Gui *et al.* (2006), depending on whether a person carries a mutation and/or a family history, we partition the entire population into 5 sub-populations as follows:

(1) sub-population 1: not mutation carriers, not having family history

(2) sub-population 2: carrying mutation BRCA1, with family members being mutation carriers

(3) sub-population 3: not carrying mutation BRCA1, with family member being mutation carriers

(4) sub-population 4: carrying mutation BRCA2, with family members being mutation carriers

(5) sub-population 5: not carrying mutation BRCA2, with family members being mutation carriers.

The mutation frequencies of these sub-populations were given as in Section 3.5. The onset rates of a family history relevant to OC for each sub-populations are shown in Figure 7.7.

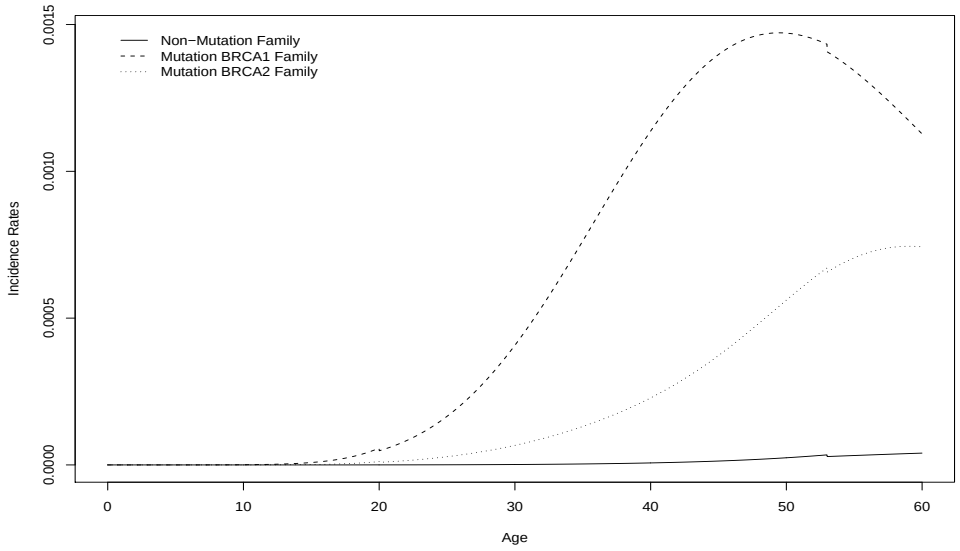

Figure 7.7: Rates of onset of family history of OC in risk sub-populations.

7.2 Application of the Life History Model to IPI Business

7.2.1 Numerical Procedures

Following the notation in Lu *et al.* (2008), we define $\mathbb{H}$ to be the set of all healthy states, including 'Normal' and 'Able to Work' states, $\mathbb{S}$ to be the set 'Other Sickness' states, and $\mathbb{O}$ to be the set of OC states (including recurrences). Suppose an applicant in subpopulation i, insured at age x, starting in state $i0$, is in state ij at age $x+t$ and has been in that state for duration z. A continuous premium at annual rate a^{ij}_{x+t} would be paid if $ij \in \mathbb{H}$ or $ij \in \mathbb{S} \cup \mathbb{O}$ with $z \leq d$; and a continuous benefit at rate b^{ij}_{x+t} will be claimed if $ij \in \mathbb{S} \cup \mathbb{O}$ with $z > d$.

In Chapter 2, we already introduced an idea to bring a semi-Markov model back to a Markov model regime by assuming that on entry into a state from which the probability of exit is duration-dependent future premiums cease and a sum assured of ${}_{0,t}{}^{i}V^{j}_{x}$ is paid, where ${}_{z,t}{}^{i}V^{j}_{x}$ is the prospective reserve for a life in state ij at age $x+t$ with duration z since entry into the state. When we apply this technique to the model in Figure 6.2, it requires more work because duration-dependent decrements occur not only into terminal states, but throughout the whole model. So we need to calculate ${}_{0,t}{}^{i}V^{j}_{x}$ for each transition into a duration-dependent state, starting from terminal states and progressing backward until all dependence on duration is eliminated. Ultimately, all expected prospective payments will be represented as the propective reserves ${}_{0,0}{}^{i}V^{0}_{x}$ in the starting states $i0$ at time 0. This is equivalent to to insurers immediately reinsuring the policy just after issuing it. This method is called the 'backward integration method', adopted from Lu *et al.* (2008). For each state ij, let $\mathbb{I}^{ij}$ be the set of states into which there is a transition from state ij. These are called inferior states of ij; and ij is called a superior state to states in $\mathbb{I}^{ij}$.

Then the reserves can be calculated as follows:

(a) If $ij \in \mathbb{O}$,

$$
\begin{aligned}
{}_{0,t}{}^{i}V_x^j &= \int_d^{TE} e^{-\delta s} \cdot {}_s{}^{i}\overline{p}_{x+t,0}^{jj} \cdot b_{x+t+s}^{ij} ds - \int_0^d e^{-\delta s} \cdot {}_s{}^{i}\overline{p}_{x+t,0}^{jj} \cdot a_{x+t+s}^{ij} ds \\
&\quad + \sum_{ik \in \mathbb{I}^{ij}} \int_0^{TE} e^{-\delta s} \cdot {}_s{}^{i}\overline{p}_{x+t,0}^{jj} \cdot {}^{i}\mu_{x+t+s,s}^{j,k} \cdot {}_{0,t+s}{}^{i}V_x^k ds, \qquad (7.22)
\end{aligned}
$$

where:

(1) d represents the DP in years, and TE denotes the time to policy expiry, that is:

$$TE = \min(\text{Policy Term} - t, 65 - x - t)$$

and;

(2) ${}_t{}^{i}\overline{p}_{x,z}^{jj}$ is the probability that a person in state ij at age x with duration z will remain in state ij for time t, at which time the duration must be $z + t$, and:

$${}_t^{i}\overline{p}_{x,z}^{jj} = \exp\left(- \int_0^t \sum_{ik \in \mathbb{I}^{ij}} {}^{i}\mu_{x+s,z+s}^{j,k} ds \right). \qquad (7.23)$$

(b) If $ij \in \mathbb{H}$,

$$
\begin{aligned}
{}_{0,t}{}^{i}V_x^j &= \int_0^{TE} e^{-\delta s} \cdot \sum_{ik \in \mathbb{I}^{ij} \cap \mathbb{S}} \left({}_s{}^{i}p_{x+t,0}^{jk} - {}_{d,s}{}^{i}p_{x+t,0}^{jk} \right) \cdot b_{x+t+s}^{ik} \cdot r_d(x+t+s) ds \\
&\quad - \int_0^{TE} e^{-\delta s} \cdot \sum_{ik \in \mathbb{I}^{ij} \cap \mathbb{S}} \left({}_s{}^{i}p_{x+t,0}^{jk} - {}_{d,s}{}^{i}p_{x+t,0}^{jk} \right) \cdot a_{x+t+s}^{ik} \cdot (1 - r_d(x+t+s))\, ds \\
&\quad - \int_0^{TE} e^{-\delta s} \cdot {}_s{}^{i}p_{x+t,0}^{jj} \cdot a_{x+t+s}^{ij} ds \\
&\quad - \int_0^{TE} e^{-\delta s} \cdot \sum_{ik \in \mathbb{I}^{ij} \cap \mathbb{S}} \left({}_{d,s}{}^{i}p_{x+t,0}^{jk} \cdot a_{x+t+s}^{ik} \right) ds \\
&\quad + \sum_{ik \in \mathbb{I}^{ij} \cap \mathbb{O}} \int_0^{TE} e^{-\delta s} \cdot {}_s{}^{i}p_{x+t,0}^{jj} \cdot {}^{i}\mu_{x+t+s,s}^{j,k} \cdot {}_{0,t+s}{}^{i}V_x^k ds, \qquad (7.24)
\end{aligned}
$$

where ${}_{w,t}{}^{i}p_{x,z}^{jk}$ is the probability that a person in state ij at age x with duration z will be in state ik at age $x+t$, with duration $z+t \le w$ (note that when $w \ge z+t$, ${}_{w,t}{}^{i}p_{x,z}^{jk} = {}_{t}{}^{i}p_{x,z}^{jk}$). It can be shown that ${}_{t}{}^{i}p_{x,z}^{jk}$ and ${}_{w,t}{}^{i}p_{x,z}^{jk}$ satisfy the following differential equations:

$$
\begin{aligned}
\frac{\partial}{\partial t}{}_{t}^{i}p_{x,z}^{jj} &= \int_0^t {}_s{}^{i}p_{x,z}^{jj} \cdot \sum_{ik \in \mathbb{I} \cap \mathbb{S}} \left({}^{i}\mu_{x+s,z+s}^{j,k} \cdot {}_{t-s}{}^{i}p_{x+s,0}^{\overline{kk}} \cdot {}^{i}\mu_{x+t,t-s}^{k,j} \right) ds \\
&\quad - {}_{t}^{i}p_{x,z}^{jj} \cdot \sum_{ik \in \mathbb{I}^{ij}} {}^{i}\mu_{x+t,z+t}^{j,k} \qquad (7.25)
\end{aligned}
$$

$$
\frac{\partial}{\partial t}{}_{w,t}{}^{i}p_{x,z}^{jk} = \begin{cases} {}_{t-w}{}^{i}p_{x,z}^{jj} \cdot \sum_{ik \in \mathbb{I}^{ij}} \left({}^{i}\mu_{x+t-w}^{j,k} \cdot {}_{w}{}^{i}p_{x+t-w,0}^{\overline{kk}} \right) & w < t \\ 0 & w \ge t. \end{cases} \qquad (7.26)
$$

We have the following comments:

(a) ${}^{i}\mu_{x,z}^{j,k}$ can be written ${}^{i}\mu_{x}^{j,k}$ or ${}^{i}\mu_{z}^{j,k}$ if it depends only on current age x or duration z in the current state, respectively.

(b) Equations (7.22) and (7.24) are recursive and for each state ij, ${}_{0,t}{}^{i}V_x^j$ can be calculated once ${}_{0,t+s}{}^{i}V_x^k$ has been obtained for all its inferior states for any further time beyond $x+t$. The terminal state is 'Dead' and has no inferior states.

(c) Equations (7.22) – (7.24) are solved by numerical integration using Simpson's rule with stepsize 1/156 years (implying three steps in one week) and $\delta = 0.05$; Equations (7.25) and (7.26) are solved by a fourth-order Runge-Kutta method with step-size 0.0005 years and initial conditions ${}_{0}{}^{i}p_{x,z}^{jj} = 1$ and ${}_{0,t}{}^{i}p_{x,z}^{jk} = 0$, for all t. Please see Appendix D for technical details about fourth-order Runge-Kutta method.

(d) In Equation (7.24), for each state $ij \in \mathbb{H}$, by defining the probabilities ${}_{t}{}^{i}p_{x,z}^{jj}$ and ${}_{w,t}{}^{i}p_{x,z}^{jk}$, we can discount all prospective cashflows in every state $ik \in \mathbb{I}^{ij} \cap \mathbb{S}$ back to the time of entering state ij. In this way, healthy states and their corresponding 'Other Sickness' states are treated as a unit and hence the periodic transitions between them are eliminated.

(e) We should notice that the use of $r_d(x+t+s)$ in Equation (7.24) is an approximation, which arises in the following way. It is impractical to compute the EPV of benefit cashflows by integrating over the joint density of the times of transition, because repeated transitions mean that the number of transitions is not bounded. That is, we would require an infinite-dimentional multiple integral (although truncation at some point would presumably give satisfactory accuracy). Therefore we write the EPV as a single integral over the probability of the 'status' — ill for longer than the DP d year — that causes the benefit to be paid. However, inspection of Equation (7.8) shows that the function $r_d(x+t+s)$ is, in fact, a function of the age at which a claim (equivalently, spell of illness) commences, although the notation suppresses that fact. Therefore the present value of benefit payable at time $x+t+s$ should be integrated over the density of the duration of the claim at that time. From CMIB 13

(1993), the claim ratios are not very dependent on age, so we avoid this additional complexity by using the claim ratio at age $x + t + s$ and ignoring current duration.

7.2.2 Premium Ratings for the General Population

The EPV of the unit benefit can be obtained using the equations in Section 7.2.1 with the premium rate set to zero. In Table 7.4, we show the EPVs of a unit benefit of IPI cover of £1 *per annum* for various DPs and various scenarios. In Table 7.5, we show the EPVs of an annuity of £1 *per annum* payable continuously while healthy or sick with duration less than the DP. In Table 7.6, we show the net continuous premium required. We can observe that for the same starting age and policy term, as the DPs increase, the EPVs of the benefit and level net premium decrease whereas the EPVs of the annuity increase. This is simply because shorter DPs imply earlier benefit claims.

Table 7.4: Expected present values of the IPI claim annuity of £1 *per annum* payable continuously while sick with duration exceeding the DP, according to the standard female experience for 1995–1998 and based on the model in Figure 6.2.

		Deferred Period			
Age	Term	DP1	DP4	DP13	DP26
20	10	0.1608	0.0183	0.0115	0.0086
	20	0.4171	0.0629	0.0423	0.0333
	30	0.7327	0.1347	0.0937	0.0806
	45	1.2279	0.3069	0.2317	0.2383
30	10	0.2728	0.0472	0.0288	0.0211
	20	0.7429	0.1552	0.1048	0.0911
	35	1.7125	0.4761	0.3518	0.3834
40	10	0.4622	0.1029	0.0669	0.0613
	25	1.9421	0.5940	0.4433	0.5156
50	15	1.7213	0.5823	0.4345	0.5359

Table 7.5: Expected present values of the IPI premium annuity of £1 *per annum* payable continuously while healthy or sick with duration less than the DP, according to the standard female experience for 1995–1998 and based on the model in Figure 6.2.

		Deferred Period			
Age	Term	DP1	DP4	DP13	DP26
20	10	7.6769	7.8446	7.8503	7.8532
	20	12.0729	12.5361	12.5580	12.5682
	30	14.5049	15.2999	15.3435	15.3584
	45	15.4689	16.8741	17.0294	16.9262
30	10	7.5460	7.8123	7.8288	7.8363
	20	11.6690	12.4025	12.4521	12.4638
	35	14.1712	15.7072	15.8234	15.7780
40	10	7.3147	7.7299	7.7609	7.7634
	25	11.7055	13.3022	13.4298	13.3312
50	15	8.3684	9.6592	9.7750	9.6478

Table 7.6: Level net premium for an IPI benefit of £1 *per annum*, according to the standard female experience for 1995–1998 and based on the model in Figure 6.2.

		Deferred Period			
Age	Term	DP1	DP4	DP13	DP26
20	10	0.0209	0.0023	0.0015	0.0011
	20	0.0346	0.0050	0.0034	0.0026
	30	0.0505	0.0088	0.0061	0.0052
	45	0.0794	0.0182	0.0136	0.0141
30	10	0.0362	0.0060	0.0037	0.0027
	20	0.0637	0.0125	0.0084	0.0073
	35	0.1208	0.0303	0.0222	0.0243
40	10	0.0632	0.0133	0.0086	0.0079
	25	0.1659	0.0447	0.0330	0.0387
50	15	0.2057	0.0603	0.0444	0.0555

7.2.3 Premium Ratings for Mutation Carriers

In Table 7.7 we show level premium rates both as absolute values and as a percentage of the standard rates, for women with a known BRCA1/2 mutation, and we make the following comments:

(a) A BRCA1 mutation is much more severe than a BRCA2 mutation. Extra premiums for all BRCA2 cases are less than 50%, which is much smaller than the corresponding BRCA1 case.

(b) Insurability for BRCA1 mutation carriers varies between different DPs. Most Cases with DP 1 week are insurable. For cases with DP 4 and 13 weeks, all BRCA1 mutation carriers should be insurable at extra premium rates. However for cases with DP 26 weeks, BRCA1 mutation carriers at low ages (20 or 30 year old) buying shorter term IPI (10 or 20 years) are generally declined; other cases might be insurable but at extra premium rates

(c) All premium increases for BRCA2 mutation carriers are below 50%, within the limit for insurability of about 250% of the standard rates. Therefore, in all cases BRCA2 mutation carriers should be insurable.

Table 7.7: Expected present values of an IPI benefit of £1 *per annum* payable continuously while sick with duration exceeding the DP for mutation BRCA1 and BRCA2 carriers, according to the standard female experience for 1995–1998 and based on the model in Figure 6.2.

Gene			Deferred Period							
Mutation	Age	Term	DP1		DP4		DP13		DP26	
BRCA1	20	10	0.0289	(138%)	0.0057	(245%)	0.0041	(279%)	0.0037	(336%)
		20	0.0460	(133%)	0.0104	(207%)	0.0075	(223%)	0.0069	(259%)
		30	0.0630	(125%)	0.0154	(174%)	0.0112	(184%)	0.0104	(199%)
		45	0.0819	(103%)	0.0202	(111%)	0.0152	(112%)	0.0154	(109%)
	30	10	0.0548	(151%)	0.0146	(241%)	0.0103	(281%)	0.0095	(355%)
		20	0.0814	(128%)	0.0221	(177%)	0.0160	(190%)	0.0150	(206%)
		35	0.1225	(101%)	0.0329	(108%)	0.0243	(109%)	0.0257	(106%)
	40	10	0.0869	(138%)	0.0263	(198%)	0.0188	(218%)	0.0181	(230%)
		25	0.1672	(101%)	0.0480	(107%)	0.0357	(108%)	0.0402	(104%)
	50	15	0.2082	(101%)	0.0647	(107%)	0.0479	(108%)	0.0574	(103%)
BRCA2	20	10	0.0209	(100%)	0.0023	(100%)	0.0015	(100%)	0.0011	(100%)
		20	0.0346	(100%)	0.0050	(100%)	0.0034	(100%)	0.0027	(101%)
		30	0.0513	(102%)	0.0093	(106%)	0.0065	(106%)	0.0056	(107%)
		45	0.0796	(100%)	0.0184	(101%)	0.0138	(101%)	0.0142	(101%)
	30	10	0.0363	(100%)	0.0061	(101%)	0.0037	(101%)	0.0027	(102%)
		20	0.0652	(102%)	0.0135	(108%)	0.0092	(109%)	0.0081	(110%)
		35	0.1209	(100%)	0.0307	(101%)	0.0226	(102%)	0.0245	(101%)
	40	10	0.0674	(107%)	0.0157	(118%)	0.0105	(122%)	0.0098	(124%)
		25	0.1662	(100%)	0.0455	(102%)	0.0337	(102%)	0.0390	(101%)
	50	15	0.2065	(100%)	0.0618	(103%)	0.0456	(103%)	0.0562	(101%)

7.2.4 Premium Ratings for Women With a Family History

When genetic test results are known to insurers, the underwriting is fairly straightforward as shown in Section 7.2.3. However, genetic test results are not allowed to be used for underwriting in most countries. Therefore, family history is probably the most important information that may be used in the underwriting. We calculated the onset rate of a family history relevant to OC in Section 7.1.3. Given a family history, a level premium rate is derived, in which the EPVs of benefit and payment are both weighted by the probability of a person aged x with a family history being in each of the sub-populations (see Section 7.1.3). The level premium rate for a woman without a family history is calculated in the same way, but the weights are changed to the probability of a person aged x without a family history being in each sub-population. Table 7.8 shows the premium rates for women with and without a family history of OC both as absolute values and as a percentage of those for standard risks. We can observe that:

(a) The premium rates for women without family histories are very close to those for non-mutation carriers. In the calculation, the weight assigned to the non-mutation carrier sub-population is very high due to the nearly zero onset rate of a family history in that sub-population.

(b) The premium increases are smaller than in the corresponding cases in Table 7.7 as expected, because of the averaging-out effect between non-mutation carriers and mutation BRCA1/2 carriers.

(c) For women with family histories, the premium increases are lower than 50% in all cases. Therefore, all women with a family history of OC should be insurable if we assume that 250% of the standard rate is the threshold. Especially, in the case of DP 1 week, the premium increases are lower than 10%. This might imply that they could even buy IPI at standard rates.

Table 7.8: Level net premium for an IPI benefit of £1 *per annum* payable continuously while sick with duration exceeding the DP for women with and without a family history of BC, respectively (FH = family history present), according to the standard female experience for 1995–1998 and based on the model in Figure 6.2.

			Deferred Period							
No FH	Age	Term	DP1		DP4		DP13		DP26	
	20	10	0.0210	(100%)	0.0023	(100%)	0.0015	(100%)	0.0011	(100%)
		20	0.0346	(100%)	0.0050	(100%)	0.0034	(100%)	0.0027	(100%)
		30	0.0505	(100%)	0.0088	(100%)	0.0061	(100%)	0.0053	(100%)
		45	0.0794	(100%)	0.0182	(100%)	0.0136	(100%)	0.0141	(100%)
	30	10	0.0362	(100%)	0.0061	(100%)	0.0037	(100%)	0.0027	(100%)
		20	0.0637	(100%)	0.0125	(100%)	0.0084	(100%)	0.0073	(100%)
		35	0.1208	(100%)	0.0303	(100%)	0.0222	(100%)	0.0243	(100%)
	40	10	0.0632	(100%)	0.0133	(100%)	0.0086	(100%)	0.0079	(100%)
		25	0.1659	(100%)	0.0447	(100%)	0.0330	(100%)	0.0387	(100%)
	50	15	0.2057	(100%)	0.0603	(100%)	0.0445	(100%)	0.0555	(100%)
FH	20	10	0.0221	(106%)	0.0029	(122%)	0.0019	(127%)	0.0015	(136%)
		20	0.0363	(105%)	0.0058	(117%)	0.0040	(119%)	0.0033	(125%)
		30	0.0525	(104%)	0.0099	(112%)	0.0069	(113%)	0.0061	(116%)
		45	0.0798	(100%)	0.0185	(102%)	0.0139	(102%)	0.0143	(101%)
	30	10	0.0397	(110%)	0.0077	(127%)	0.0049	(135%)	0.0040	(149%)
		20	0.0671	(105%)	0.0144	(115%)	0.0099	(117%)	0.0088	(120%)
		35	0.1211	(100%)	0.0308	(101%)	0.0226	(102%)	0.0245	(101%)
	40	10	0.0673	(106%)	0.0156	(117%)	0.0104	(121%)	0.0097	(123%)
		25	0.1661	(100%)	0.0452	(101%)	0.0334	(101%)	0.0389	(101%)
	50	15	0.2059	(100%)	0.0607	(101%)	0.0448	(101%)	0.0557	(100%)

7.2.5 The Effect of Reduced Onset Rates

In the model in Figure 6.2, the detection rates of OC at various stages for BRCA1/2 mutation carriers are based on the study by Antoniou *et al.* (2001). Therefore, the potential for ascertainment bias is still present, and we allow for possible bias by considering incidence rates of OC for BRCA1/2 mutation carriers 50% of those based directly on Antoniou *et al.* (2001). Tables 7.9 and 7.10 show level net premiums expressed as a percentage of the standard rates for mutation carriers and for women with a family history, respectively, with incidence rates of BC for mutation carriers reduced 50% of those estimated from Antoniou *et al.* (2001). We note that:

(a) With the OC incidence rate reduced by 50%, it is not surprising to see that the corresponding premium rates are reduced as well. This would make all women with mutation BRCA1 insurable, especially those with longer DPs. Women with mutation BRCA2 are all insurable even before the incidence rate is reduced, so they are more clearly insurable. The reduced rate of onset will allow more women with mutation BRCA2 to be covered at the standard rate.

(b) Premium rates are also reduced for women with a family history. In most cases, except for policies with shorter terms or longer DPs, the premium increases are lower than 10%. These cases might be offered stardard rates.

Table 7.9: Expected present values of an IPI benefit of £1 *per annum* payable continuously while sick with duration exceeding the DP for mutations BRCA1 and BRCA2 carriers, according to the standard female experience for 1995–1998 and based on the model in Figure 6.2. Excess OC indidence rates are 50% of those observed.

Gene			Deferred Period							
Mutation	Age	Term	DP1		DP4		DP13		DP26	
BRCA1	20	10	0.0250	(119%)	0.0041	(175%)	0.0028	(192%)	0.0024	(221%)
		20	0.0405	(117%)	0.0078	(156%)	0.0055	(165%)	0.0049	(183%)
		30	0.0569	(113%)	0.0122	(139%)	0.0088	(144%)	0.0080	(152%)
		45	0.0795	(100%)	0.0188	(103%)	0.0140	(103%)	0.0143	(101%)
	30	10	0.0461	(128%)	0.0106	(176%)	0.0073	(198%)	0.0064	(237%)
		20	0.0733	(115%)	0.0177	(142%)	0.0125	(149%)	0.0115	(157%)
		35	0.1201	(100%)	0.0310	(102%)	0.0228	(103%)	0.0244	(100%)
	40	10	0.0763	(121%)	0.0205	(154%)	0.0143	(165%)	0.0135	(172%)
		25	0.1656	(100%)	0.0459	(103%)	0.0340	(103%)	0.0390	(101%)
	50	15	0.2074	(101%)	0.0626	(104%)	0.0463	(104%)	0.0567	(102%)
BRCA2	20	10	0.0209	(100%)	0.0023	(100%)	0.0015	(100%)	0.0011	(100%)
		20	0.0346	(100%)	0.0050	(100%)	0.0034	(100%)	0.0027	(100%)
		30	0.0509	(101%)	0.0090	(103%)	0.0063	(103%)	0.0054	(103%)
		45	0.0793	(100%)	0.0182	(100%)	0.0136	(100%)	0.0140	(100%)
	30	10	0.0362	(100%)	0.0061	(100%)	0.0037	(100%)	0.0027	(100%)
		20	0.0644	(101%)	0.0130	(104%)	0.0088	(104%)	0.0077	(105%)
		35	0.1205	(100%)	0.0303	(100%)	0.0223	(100%)	0.0242	(100%)
	40	10	0.0652	(103%)	0.0145	(109%)	0.0095	(111%)	0.0088	(111%)
		25	0.1658	(100%)	0.0449	(101%)	0.0332	(101%)	0.0387	(100%)
	50	15	0.2062	(100%)	0.0610	(101%)	0.0450	(101%)	0.0559	(101%)

Table 7.10: Level net premium for an IPI benefit of £1 *per annum* payable continuously while sick with duration exceeding the DP for women with and without a family history of BC, respectively (FH = family history persent), according to the standard female experience for 1995–1998 and based on the model in Figure 6.2. Excess OC incidence rates are 50% of those observed.

			Deferred Period							
No FH	Age	Term	DP1		DP4		DP13		DP26	
	20	10	0.0210	(100%)	0.0023	(100%)	0.0015	(100%)	0.0011	(100%)
		20	0.0346	(100%)	0.0050	(100%)	0.0034	(100%)	0.0027	(100%)
		30	0.0505	(100%)	0.0088	(100%)	0.0061	(100%)	0.0053	(100%)
		45	0.0794	(100%)	0.0182	(100%)	0.0136	(100%)	0.0141	(100%)
	30	10	0.0362	(100%)	0.0060	(100%)	0.0037	(100%)	0.0027	(100%)
		20	0.0637	(100%)	0.0125	(100%)	0.0084	(100%)	0.0073	(100%)
		35	0.1208	(100%)	0.0303	(100%)	0.0222	(100%)	0.0243	(100%)
	40	10	0.0632	(100%)	0.0133	(100%)	0.0086	(100%)	0.0079	(100%)
		25	0.1659	(100%)	0.0447	(100%)	0.0330	(100%)	0.0387	(100%)
	50	15	0.2057	(100%)	0.0603	(100%)	0.0444	(100%)	0.0555	(100%)
FH	20	10	0.0215	(103%)	0.0026	(111%)	0.0017	(113%)	0.0013	(118%)
		20	0.0354	(103%)	0.0054	(108%)	0.0037	(110%)	0.0030	(112%)
		30	0.0515	(102%)	0.0093	(106%)	0.0065	(107%)	0.0057	(108%)
		45	0.0794	(100%)	0.0183	(100%)	0.0137	(100%)	0.0141	(100%)
	30	10	0.0379	(105%)	0.0069	(113%)	0.0043	(117%)	0.0033	(124%)
		20	0.0654	(103%)	0.0134	(107%)	0.0092	(109%)	0.0081	(110%)
		35	0.1207	(100%)	0.0304	(100%)	0.0223	(100%)	0.0243	(100%)
	40	10	0.0653	(103%)	0.0145	(109%)	0.0095	(110%)	0.0088	(111%)
		25	0.1659	(100%)	0.0448	(100%)	0.0332	(100%)	0.0387	(100%)
	50	15	0.2059	(100%)	0.0605	(100%)	0.0446	(100%)	0.0557	(100%)

7.3 The Cost of Adverse Selection in an IPI Market

7.3.1 The Life History Model of a Women at Risk of Breast Cancer in an IPI Market

As in Chapters 2 and 4, we use CI and life insurance market models to depict the applicants' purchasing behaviour, and calculate the cost of adverse selection. In this section, we introduce an IPI market model as shown in Figure 7.8, which incorporates the insurance-purchasing behaviour of applicants at various levels of risk.

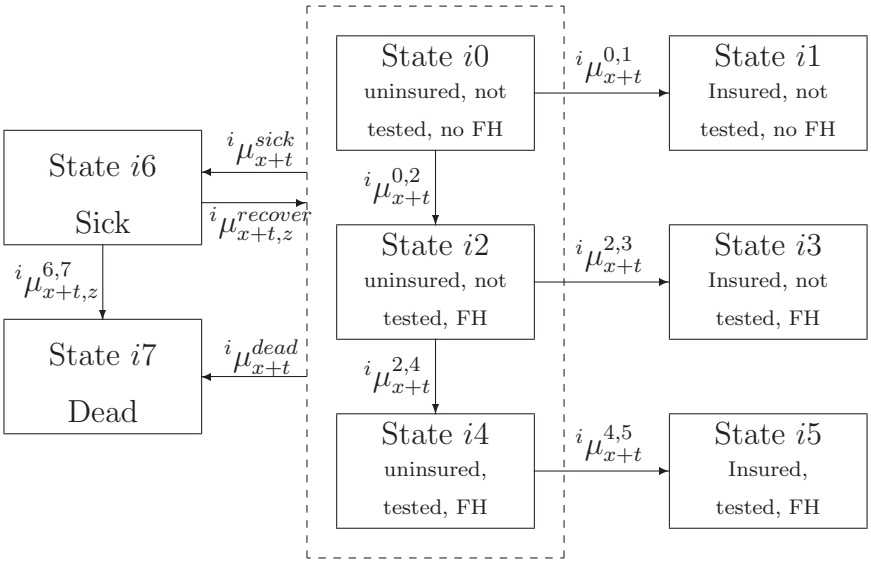

Figure 7.8: A semi-Markov model of family history, genetic testing, and IPI purchase for a person in the i^{th} risk subpopulation (FH = family history present).

Please see Lu *et al.* (2008) for model features, parameterization and numerical procedures. We briefly describe some important features as follows:

(a) Depending on whether a person carries a mutation and/or has a family history, we partition the entire population into 5 sub-populations as follows:

(1) sub-population 1: not mutation carriers, not having a family history

(2) sub-population 2: carrying mutation BRCA1, with family members being mutation carriers

(3) sub-population 3: not carrying mutation BRCA1, with family member being mutation carriers

(4) sub-population 4: carrying mutation BRCA2, with family members being mutation carriers

(5) sub-population 5: not carrying mutation BRCA2, with family members being mutation carriers.

The frequencies of these sub-populations were introduced in Section 3.5.

(b) The purpose of this model is to calculate the cost of adverse selection arising in an IPI market.

(c) We assume that an individual may fall sick or die independently of buying an IPI policy. All states in the dashed box, e.g. State $i0$, $i2$ and $i4$, have access to States $i6$ and $i7$. If persons in States $i0$, $i2$ and $i4$ enter into State $i7$ before they buy insurance, i.e. entering into State $i1$, $i3$ and $i5$, they become uninsurable. After buying an IPI policy, a woman will follow the life history model in Figure 6.2.

(d) In the model in Figure 6.2, the intensities ${}^{i}\mu_{x+t}^{sick}$, ${}^{i}\mu_{x+t}^{recover}$ and ${}^{i}\mu_{x+t,z}^{6,7}$ correspond to intensities in the basic IPI model in Figure 6.1. However, we should note that these

events occur before the individual becomes insured. Since direct estimates have not been reported yet, we use the estimates described in Section 7.1 with DP 1 week, assuming no significant difference would be expected.

(e) Intensity ${}^{i}\mu_{x+t}^{0,2}$ represents the onset rate of developing a family history in sub-population i. These intensities were introduced in Section 3.5. Intensity ${}^{i}\mu_{x+t}^{2,4}$ represents the genetic testing rate. In this thesis, we take genetic testing rate to be 0.014 per annum between ages 20 and 40 as our baseline, implying about 10% of people will be tested in 8 years. We compare this baseline with another two scenarios, namely 0.014 per annum between ages 20 and 60, and 0.035 per annum between ages 20 and 40.

(f) Persons in States $i0$, $i2$ and $i4$ have not purchased any insurance yet, while persons in States $i1$, $i3$ and $i5$ have purchased insurance. The intensities ${}^{i}\mu_{x+t}^{0,1}$, ${}^{i}\mu_{x+t}^{2,3}$ and ${}^{i}\mu_{x+t}^{4,5}$ are the annualized purchase rates. The meaning of the purchase rate is two-fold. Firstly, it represents the size of the IPI market. A purchase rate of 0.05 per annum represents what we call a large market. A purchase rate of 0.01 per annum represents what we call a small market. These are assumed to apply in sub-population 1. We call these purchase rates the 'normal' purchase rates. In sub-populations 2 – 5, persons in State $i0$ buy insurance at the normal rate in all circumstances, because these persons will be treated by insurance companies as standard risk. In the large market, persons, who have a family history are assumed to have three choices of their purchasing behaviour: normal rate, half of normal rate, or nil, meaning they do not purchase insurance at all. In a small market, these persons are assumed not to purchase insurance at all.

(g) A moratorium forbids insurance companies to use genetic test results in underwriting. Three types of moratoria are considered here, moratoria on using all genetic

test results, moratoria on using adverse genetic test results, and moratoria on both family history and genetic test results. Normally only insurance applicants with similar level of risks will be pooled into the same underwriting class, and people in the same underwriting class should be charged the same rate of premium. However, a moratorium may force underwriters to pool very different risks into the same underwriting class. A moratorium on genetic testing results alone will partition the population into two underwriting classes; family history class and non-family history class. A moratorium on genetic testing results and family history will put the entire population into a single underwriting class.

(h) Adverse selection is represented by increasing the insurance purchasing rate of persons with adverse genetic information when a moratorium on genetic test results is imposed. Moderate adverse selection is represented by a purchasing rate double the normal rate (the normal rates are 0.05 per annum in a large market and 0.01 per annum in a small market), and severe adverse selection is represented by a purchasing rate of 0.25 per annum in both large and small markets.

A moratorium on genetic testing results and family history causes premiums to increase in two different ways:

(a) A new underwriting class forms, comprising persons in all sub-populations. Persons in the higher-risk sub-populations now can purchase normal amounts of insurance at ordinary rates. This will increase the premium rate. But this is not adverse selection if their behaviour is just the same as that of persons in the low-risk subpopulations.

(b) Further, insurance applicants may increase their purchase rate in reaction to the information they have and the lower premiums they have been charged. In this way, adverse selection arises.

(i) In the model in Figure 6.2, an underwriting class $\mathcal{C}$ is represented by a set of insured states. Let ${}^{i}P$ be the proportion of women in state $i0$ at age 20. Within each underwriting class, the same premium rate $\rho^{\mathcal{C}}_{20+t}$ as shown in Equation (7.27) should be charged to cover the average risk using the Principal of Equivalence.

$$\rho^{\mathcal{C}}_{20+t} = \frac{\displaystyle\sum_{j=1,3,5}\left(\sum_{ij\in\mathcal{C}} {}^{i}P \int_0^t {}_{s}^{i}p_{20}^{0,(j-1)} \cdot {}^{i}\mu_{20+s}^{(j-1),j} \cdot {}_{t-s}{}^{i}p_{20+s}^{\overline{jj}} \cdot {}^{i}\mathbf{P}^{j}_{20+s:\overline{45-s|}} ds\right)}{\displaystyle\sum_{j=1,3,5}\left(\sum_{ij\in\mathcal{C}} {}^{i}P \int_0^t {}_{s}^{i}p_{20}^{0,(j-1)} \cdot {}^{i}\mu_{20+s}^{(j-1),j} \cdot {}_{t-s}{}^{i}p_{20+s}^{\overline{jj}} ds\right)} \tag{7.27}$$

In Equation (7.27), let ${}_{t}^{i}p_{x}^{jj}$ be the probability that a person in state ij at age x will be in the same state at age $x+t$; and ${}_{t}^{i}p_{x}^{\overline{jj}}$ be the probability that a person in state ij at age x will remain in state ij for time t. We define ${}^{i}\mathbf{P}^{j}_{x:\overline{n}|}$ as the annual net rate of premium for unit benefit for policyholders in state ij who bought insurance at age x, with policy term n. ${}^{i}\mathbf{P}^{j}_{20+s:\overline{45-s|}}$ is calculated according to the model in Figure (6.2). We explain the computation of the probabilities ${}_{t}^{i}p_{x}^{\overline{jj}}$ and ${}_{t}^{i}p_{x}^{k,(j-1)}$ as follows:

(a) The probability ${}_{t}^{i}p_{x}^{\overline{jj}}$ is computed using equations analogous to Equation (7.23).

(b) If $j = 1$, the probability ${}_{t}^{i}p_{x}^{0,(j-1)}$ is computed using equations analogous to Equation (7.24).

(c) If $j = 3$, the probability ${}_{t}^{i}p_{x}^{0,(j-1)}$ is computed as:

$${}_{t}^{i}p_{x}^{0,2} = \int_0^t {}_{s}^{i}p_{x}^{0,0} \cdot {}^{i}\mu_{20+s}^{0,2} \cdot {}_{t-s}{}^{i}p_{20+s}^{\overline{22}} ds. \tag{7.28}$$

(d) If $j = 5$, the probability ${}_{t}^{i}p_{x}^{0,(j-1)}$ is computed as:

$${}_{t}{}^{i}p_{x}^{0,4} = \int_0^t {}_{s}^{i}p_{x}^{0,2} \cdot {}^{i}\mu_{20+s}^{2,4} \cdot {}_{t-s}{}^{i}p_{20+s}^{\overline{44}} ds. \tag{7.29}$$

(j) With these rates of premium, policy values conditional on being in any state are obtained using the 'backward integration method' as in Equation (7.22) and (7.24). Then the expected loss in the entire market is the weighted average of these policy values in each starting state, the weights being the corresponding proportions at outset. If insurance is purchased exactly as assumed when calculating $\rho^{\mathcal{C}}_{x+t}$, there is no adverse selection and the expected loss is zero. However, when there exists a moratorium, adverse selection may occur, and more insurance might be purchased than assumed. If we keep charging the premium rate $\rho^{\mathcal{C}}_{x+t}$, a loss emerges, which is the cost of adverse selection. In order to recoup the loss, insurers would have to increase all premiums by:

$$\frac{\text{EPV of loss with adverse selection} - \text{EPV of loss without adverse selection}}{\text{EPV of premiums payable with adverse selection}}. \tag{7.30}$$

This quantity is our measure of the level of adverse selection.

7.3.2 Moratoria on Using Genetic Test Results

When there is a moratorium on genetic test results alone, either on all genetic test results or adverse test results, it will partition the population into two underwriting classes: one with persons not at risk, and being charged the ordinary rate, and another with family history and being charged a higher premium rate. Tables 7.11 – 7.14 show the percentage premium increases required to cover the cost of adverse selection, under moratoria on all genetic test results or adverse test results respectively, with the baseline genetic test rate (0.014 per annum between ages 20 to 40), for IPI markets with different DPs. The results are fairly small, even negligible if we use 0.1% as the threshold, however we observe the following:

(a) Under the same circumstances, percentage premium increases are bigger with longer DPs.

(b) The percentage premium increases are bigger in small market as expected.

(c) The costs of adverse selection are slightly higher under a moratorium on all genetic test results than that under a moratorium on adverse genetic test results. This is because tested persons who are not mutation carriers will be "exonerated" and charged standard premiums.

(d) The impact on an IPI market is much smaller, compared with that on a life or CI insurance market. The absolute losses caused by adverse selection in these three markets are roughly of the same magnitude. However, the EPVs of the premiums payable for unit benefit in an IPI market are far larger than the other two markets.

Table 7.15 and 7.16 show the percentage premium increases required to recoup the cost of severe adverse selection, under moratoria on all genetic test results or adverse test results, with a rate of genetic testing of 0.014 per annum between ages 20 to 60 or 0.035 per annum between ages 20 to 40, for an IPI market with different DPs. We observe that if we extend the upper age limit of genetic testing from 40 to 60, the cost of adverse selection will increase very slightly, while if we increase the genetic test rate from 0.014 per annum to 0.035 per annum, the cost of adverse selection will certainly increase as well. However, none of these two scenarios cause substantial premium increases, even if there is severe adverse selection.

7.3.3 A Moratorium on Using Family History and Genetic Test Results

We already mentioned that a moratorium on family history and genetic test results will lead to a single underwriting class, and this will cause the insurance premium to increase

Table 7.11: Percentage increases in premium rates for IPI cover with DP 1 week, under a moratorium on all genetic test results and adverse results respectively, for a market operating between ages 20 and 65.

Deferred Period	Adverse Selection	Market Size	Rate of Purchase with A Family History	Moratorium on using All test results %	Moratorium on using Adverse selection %
DP1	Severe	Large	Same as 'normal'	0.00000990	0.00000985
			Half as 'normal'	0.00001424	0.00001416
			Nil	0.00007481	0.00007443
		Small	Nil	0.00042867	0.00041153
DP1	Moderate	Large	Same as 'normal'	0.00000397	0.00000393
			Half as 'normal'	0.00000755	0.00000750
			Nil	0.00004788	0.00004768
		small	Nil	0.00007646	0.00007494

in two different ways:

(a) A new underwriting class forms, comprising persons in all sub-populations. Persons in the higher-risk sub-populations now can purchase normal amounts of insurance at ordinary rates. This will increase the premium rate. But this is not adverse selection if their behaviour is just the same as that of persons in the low-risk subpopulations.

(b) Further, insurance applicants may increase their purchase rate in reaction to the information they have and the relatively lower premiums they have been charged.

Table 7.12: Percentage increases in premium rates for IPI cover with DP 4 weeks, under a moratorium on all genetic test results and adverse results respectively, for a market operating between ages 20 and 65.

Deferred Period	Adverse Selection	Market Size	Rate of Purchase with A Family History	Moratorium on using All test results %	Moratorium on using Adverse selection %
DP4	Severe	Large	Same as 'normal'	0.00001942	0.00001934
			Half as 'normal'	0.00002939	0.00002924
			Nil	0.00012470	0.00012395
		Small	Nil	0.00070254	0.00067866
DP4	Moderate	Large	Same as 'normal'	0.00000836	0.00000828
			Half as 'normal'	0.00001658	0.00001648
			Nil	0.00008316	0.00008282
		small	Nil	0.00013465	0.00013238

In this way, adverse selection arises.

Table 7.17 shows the premium increases for different DPs. Note that since there is only one underwriting class, the purchase rates in all sub-populations are the same as normal rate (0.05 per annum in the large market and 0.01 per annum in the small market). In all cases, we take the genetic testing rate to be 0.035 per annum between ages 20 and 40. We observe the following:

(a) A new underwriting class causes much larger percentage premium increases than

Table 7.13: Percentage increases in premium rates for IPI cover with DP 13 weeks, under a moratorium on all genetic test results and adverse results respectively, for a market operating between ages 20 and 65.

Deferred Period	Adverse Selection	Market Size	Rate of Purchase with A Family History	Moratorium on using All test results %	Moratorium on using Adverse selection %
DP13	Severe	Large	Same as 'normal'	0.00002146	0.00002137
			Half as 'normal'	0.00003251	0.00003234
			Nil	0.00013202	0.00013135
		Small	Nil	0.00074403	0.00071950
DP13	Moderate	Large	Same as 'normal'	0.00000927	0.00000922
			Half as 'normal'	0.00001838	0.00001828
			Nil	0.00008835	0.00008781
		small	Nil	0.00014346	0.00014120

adverse selection. The total effect of a moratorium on genetic test results and family history certainly gives rise to much bigger costs of adverse selection than a moratorium on genetic test results alone.

(b) In a small market, the cost of adverse selection is much bigger than the corresponding case in a large market, although the absolute value is still very small.

Table 7.14: Percentage increases in premium rates for IPI cover with DP 26 weeks, under a moratorium on all genetic test results and adverse results respectively, for a market operating between ages 20 and 65.

Deferred Period	Adverse Selection	Market Size	Rate of Purchase with A Family History	Moratorium on using All test results %	Moratorium on using Adverse selection %
DP26	Severe	Large	Same as 'normal'	0.00002327	0.00002317
			Half as 'normal'	0.00003472	0.00003451
			Nil	0.00014038	0.00013939
		Small	Nil	0.00078251	0.00075750
DP26	Moderate	Large	Same as 'normal'	0.00000990	0.00000985
			Half as 'normal'	0.00001937	0.00001925
			Nil	0.00009361	0.00009304
		small	Nil	0.00014972	0.00014496

Table 7.15: Percentage increases in premium rates for IPI cover, under a moratorium on all genetic test results and adverse results respectively, under severe adverse selection, with genetic testing rate of 0.014 per annum, for a market operating between ages 20 and 65.

Deferred	Market	Rate of Purchase with	Moratorium on using	
Period	Size	A Family History	All test results	Adverse selection
		Same as 'normal'	0.00000792	0.00000787
DP1	Large	Half as 'normal'	0.00001127	0.00001121
		Nil	0.00005611	0.00005577
	Small	Nil	0.00030027	0.00029105
		Same as 'normal'	0.00001376	0.00001363
DP4	Large	Half as 'normal'	0.00002058	0.00002045
		Nil	0.00008579	0.00008544
	Small	Nil	0.00044579	0.00043208
		Same as 'normal'	0.00001515	0.00001505
DP13	Large	half as 'normal'	0.00002272	0.00002260
		Nil	0.00009016	0.00008961
	Small	Nil	0.00046715	0.00046603
		Same as 'normal'	0.00001697	0.00001686
DP26	Large	Half as 'normal'	0.00002517	0.00002501
		Nil	0.00009719	0.00009668
	Small	Nil	0.00049897	0.00049134

Table 7.16: Percentage increases in premium rates for IPI cover, under a moratorium on all genetic test results and adverse results respectively, under severe adverse selection, with genetic testing rate of 0.035 per annum, for a market operating between ages 20 and 65.

Deferred	Market	Rate of Purchase with	Moratorium on using	
Period	Size	A Family History	All test results	Adverse selection
		Same as 'normal'	0.00001879	0.00001869
DP1	Large	Half as 'normal'	0.00002667	0.00002650
		Nil	0.00013254	0.00013176
	Small	Nil	0.00071002	0.00069862
		Same as 'normal'	0.00003264	0.00003244
DP4	Large	Half as 'normal'	0.00004870	0.00004845
		Nil	0.00020251	0.00020129
	Small	Nil	0.00105382	0.00103742
		Same as 'normal'	0.00003593	0.00003571
DP13	Large	half as 'normal'	0.00005374	0.00005341
		Nil	0.00021281	0.00021153
	Small	Nil	0.00110519	0.00108869
		Same as 'normal'	0.00004021	0.00004005
DP26	Large	Half as 'normal'	0.00005939	0.00005903
		Nil	0.00022935	0.00022797
	Small	Nil	0.00118127	0.00116005

Table 7.17: Percentage increases in premium rates for IPI cover, under a moratorium on all genetic test results and family history, with genetic testing rate of 0.035 per annum between ages 20 and 40, for a market operating between ages 20 and 65.

Deferred Period	Cost From Market Size	Cost From New Underwriting Classes %	Cost From Severe Adverse Selection %	Cost From Moderate Adverse Selection %
DP1	Large	0.02780000	0.00053196	0.00022051
	Small	0.02690000	0.00461888	0.00045369
DP4	Large	0.08667000	0.00095592	0.00044080
	Small	0.08492000	0.00865876	0.00098525
DP13	Large	0.13498000	0.00103801	0.00048384
	Small	0.13023000	0.00952810	0.00109825
DP26	Large	0.14052000	0.00120828	0.00056289
	Small	0.13859000	0.01110835	0.00124785

Chapter 8

Conclusions and Further Research

8.1 Conclusions

8.1.1 The Overall Impact of Genetic Information On the Insurance Industry

In this thesis, our purpose is to assess the overall impact of genetic information on the insurance industry using the "bottom-up" approach. This work started with detailed individual studies of each genetic disorder of interest, and proceeded by aggregating all these individual studies. We chose six genetic disorders, taken to have a significant impact on insurance, including APKD, EOAD, HD, MD, HNPCC and BC & OC. All these disorders have been studied before except MD. Our work therefore started with an individual study of MD. We answered two questions: the insurability of insurance applicants, and the cost of adverse selection. We concluded as follows:

(a) Underwriting based on genetic test results will render most insurance applicants carrying mutations uninsurable, with few exceptions. For mutation CTG250+ carriers, with payment for MD arising at Stage II, all mutation carriers become uninsurable,

while for mutation CTG250– carriers, they might be insured at a high age at an increased premium rate. For both mutation CTG250+ and CTG250– carriers, with payment for MD arising at stage III, they might be insurable at a high age and at an increased premium rate. The premium rates for mutation CTG250+ carriers are much higher than that for mutation CTG250– carriers, consistent with the higher risks that mutation CTG250+ carries.

(b) Underwriting based on family history will still make most insurance applicants with family histories uninsurable, with a few exceptions appearing at high ages.

(c) The costs of adverse selection are very small, chiefly because of the low prevalence rate of mutations. In a small market the cost of adverse selection is higher than that arising in a large market. But even in a small market, if family history underwriting is still allowed, premium increases arising from adverse selection do not exceed 0.1% assuming a high rate of genetic testing and extreme adverse selection. A moratorium on both genetic test results and family history will increase premiums more noticeably, especially in a small market. However, in absolute terms, the premium increases are still small.

(d) The conclusions we draw in respect of life insurance are very similar to those in respect of CI insurance. However, under the same circumstances, the increase in life insurance premium rates is lower than in CI insurance, and the potential costs of adverse selection arising in a life insurance market are also lower than the same case in a CI insurance market.

We then reviewed five other genetic disorders in Chapter 3. In Chapter 4, we constructed the Markov or semi-Markov model, applied the epidemiological information of the relevant genetic disorders, and quantified the impact of genetic information relating to all six genetic disorders on both a CI insurance market and a life insurance market.

We then discussed the adequacy and efficacy of our model, in order to complete the "bottom-up" approach.

In an insurance market, both insurance applicants and insurance companies are concerned about genetic information. Therefore the impact of genetic information should be assessed from both insurance applicants' and insurance companies' perspectives. For the former, the question is about the insurability of insurance applicants, which has been answered in each individual study, whereas for the latter, the question is about the cost of adverse selection, which is answered in Chapter 4. We concluded as follows:

(a) Under the same circumstances, the cost of adverse selection is less in a life insurance market than in a CI insurance market. The cost of adverse selection is more noticeable in a small market than in a large market.

(b) Under a moratorium on genetic information, with moderate adverse selection, generally the cost of adverse selection is negligible (less than 0.1% of total premiums) with few exceptions, which appears in the case of the small market. However, with severe adverse selection, the costs of adverse selection become noticeable especially in the small market.

(c) The cost of adverse selection under a moratorium on adverse genetic test information causes a smaller premium increase than a moratorium on all genetic test information. The reason is because tested persons who are not mutation carriers will be "exonerated" and charged standard premiums. These people are therefore removed from the underwriting class rated for family history, which then contains a higher proportion of mutation carriers, so the premium charged in respect of this class is higher.

(d) A moratorium on both genetic test results and family history will increase the premium in two ways: consolidation of underwriting classes and adverse selection.

The premium increases due to the former are noticeable in both large and small markets. The premium increases arising from the latter are also noticeable. The highest premium increases are about 2% in a CI insurance market and 1% in a life insurance market both in the case of small market, extreme adverse selection and an exceptionally high rate of genetic testing (0.035 per annum). Insurers' chief concern is that this increase in premium rates, especially in a small market, might drive persons not at risk out of the market, cause premium rates to increase further, and lead to the collapse of the entire market.

(e) As stated above, insurers' concerns have been that the premium increases could lead to market collapse, which is called an 'adverse selection spiral' (see Section 1.2.3). Macdonald & Tapadar (2010) concluded that 'no convincing evidence that adverse selection is a serious insurance risk, even if information about multifactorial genetic disorders remains private'. Therefore, we shall believe that our results are consistent with this study and that that 'adverse selection spiral' is speculative and should not incur significant costs for the insurance companies.

(f) In the "bottom-up" approach, we selected six genetic disorders to study in this thesis. These disorders are taken to have significant impact on the insurance industry. We shall see that our selection is adequate. In terms of single-gene disorders, we have included four typical autosomal dominant single-gene disorders. In terms of single gene subsets of common disorders, we considered HNPCC, a hereditary type of CRC, which is the third most common form of cancer and second leading cause of cancer-related death; we also considered BC, the most common cancer, and OC, the second most common cancer in women. Inclusion of more disorders does not necessarily increase the cost of adverse selection, because diseases like FAP and MEN are typical examples of cancer for which genetic testing and early treatment should lead to substantially better outcomes. In many ways APKD is the least concerning

disease to be included since the insured event (ESRD) is most often predicted by ultrasound, not genetic test. Therefore, we shall trust that our selection of genetic disorders is adequate and our results are robust. By including major example of different types of single-gene disorders (the only ones for which adequate epidemiology exist) we establish a credible order of magnitude for the financial impact of all single-gene disorders, including the large number of extremely rare disorders where epidemiology will never support a trivial modelling. In this sense, we have completed the 'bottom-up' approach as far as we rationally can.

(g) In modelling the onset rates of these diseases, because of retrospective ascertainment schemes, these rates are generally over estimated, which results in overestimated premiums and potential cost of adverse selection. Even so, we find in Sections 4.2 and 4.3 that in most cases the cost of adverse selection is not noticeable, generally because of the low prevalence of genetic disorders. Finding the exact answer to the cost of adverse selection is not an easy job. We might reasonably take the results presented in Chapter 4 as the upper bounds for measuring the cost of adverse selection, so that we shall know the magnitude of genetic risk to insurers. Macdonald (2003b) concluded that the increase in life insurance premium rates should be under 10% with a moratorium on using genetic testing results and family history, using the "top-down" approach with very broad assumptions about the genetic morbidity. In this thesis, we narrowed down this range to 1% in the life insurance market. Therefore, we conclude that our results are consistent with other conclusions in other studies and are reliable upper bounds for the cost of adverse selection.

(h) The genetic risk might cause concerns for insurers. However, we should not forget the other factors, which life insurers benefit from, e.g. development of health care and the general trend of mortality improvement. These benefits greatly overwhelm the genetic risk. Therefore, in total, genetic risk should not have significant impact

on the insurance industry.

(i) An interesting question to ask is that whether the marginal costs of genetic disorders are sub-additive, additive or super-additive, as opposed to the overall cost of genetic information. The question is hard to answer for the following reasons:

   (a) The common assumption about the purchase rate is consistent in all individual studies. However, various assumptions about genetic testing results were used in the calculation. This has some justification from studies of genetic testing.

   (b) An individual study of HNPCC in life insurance market has not been done yet. Therefore it is hard to answer this question in life insurance market.

   We summarise the cost of adverse selection for males and females from relevant individual studies in Table 8.1 and compare the sum of these marginal effect with the aggregate result presented in Chapter 4. The common assumptions for these numerical results are: (1) large market; (2) normal insurance purchase rate for at-risk persons; (3) severe adverse selection; (4) moratoria on using adverse test results. The assumptions about the genetic testing rate varies and are listed in Table 8.1.

   We can see some features in Table 8.1:

   (a) Based on these results, we can see that the sum of marginal costs are 0.039% and 0.035 for females and males respectively, whereas the aggregate results are 0.040% and 0.038% for females and males respectively.

   (b) Although the sum of marginal costs are slightly lower than the aggregate cost, the difference is small. This difference might be very sensitive to the changes to genetic testing assumptions in the case of EOAD, HNPCC and BC & OC.

   (c) Although we can not make definitive conclusion at this stage, we should expect that the marginal costs should be almost additive, considering the fact that:

Table 8.1: A table listing the marginal cost of adverse selection of six genetic disorders and comparing the sum of marginal costs with the aggregate cost. The common assumptions for these numerical results are: (1) large market; (2) normal insurance purchase rate for at-risk persons; (3) severe adverse selection; (4) moratoria on using adverse test results.

		Marginal Costs of Adverse Selection	
Genetic Disorders	Genetic Testing Rate	Females	Males
		%	%
APKD	0.035 per annum between ages 20 and 40	0.018	0.017
EOAD	0.01 per annum between ages 20 and 60	0.004	0.004
HD	0.035 per annum between ages 20 and 40	0.010	0.010
MD	0.035 per annum between ages 20 and 40	0.005	0.004
HNPCC	0.1 per annum between ages 20 and 60	0.001	0.000
BC & OC	0.04 per annum between ages 20 and 60	0.001	0.000
	Sum of the marginal costs	0.039	0.035
Overall Costs	0.035 per annum between ages 20 and 40	0.040	0.038

(1) it is very unlikely that one individual are at risk of more than one single-gene disorders.

(2) it is rare that an individual at risk of single-gene disorders develops a family history of multifactorial disorders, e.g. breast cancers, and selects against insurers.

8.1.2 CRC and CRC Screening Program

In Chapter 5, we studied the reduction effect of CRC screening programs on the risks associated with CRC, e.g. the onset rates and mortality caused by CRC, and consequently

the effect on life insurance for both the general population and persons at risk of HNPCC. We concluded as follows:

(a) The Bowel Screening Program has been proved to reduce mortality caused by CRC. A meta-analysis, Hewitson *et al.* (2007), showed that participants allocated to screening had a 16% reduction in the relative risk (RR) of CRC mortality (RR 0.84, CI: 0.78-0.90).

(b) The CRC Surveillance Program also reduced significantly the onset rate of CRC for persons at risk of HNPCC. Järvinen *et al.* (2000) concluded that in mutation-positive subjects alone, the onset rate of CRC was reduced by 56%.

(c) Reduction in population mortality attributable to the Bowel Screening Program leads to a decrease in life insurance premium rates as expected. The most significant fall in premium rate (2%) appears in the case of males aged 50 seeking 10 years of cover.

(d) With underwriting based on genetic test results, life insurance premium rates for mutation carriers should be reduced if they choose to take part in a CRC Surveillance Program. Most cases are insurable at increased premium rates, while female MSH2 mutation carriers are most likely to be declined with their premium rates being between 300% – 400% times that for standard risks.

(e) Family history underwriting averages out the premium rates for mutation and non-mutation carriers, and hence we can offer lower rates for persons carrying a family history. Since most cases are insurable at increased premium rates with underwriting based on genetic test results, family history underwriting will make all cases insurable at lower increased premiums.

8.1.3 Ovarian Cancer and IPI

In Chapters 6 and 7, we studied the impact of genetic information relevant to OC on the IPI market. We concluded as follows:

(a) The relative increase in premium rates for IPI are lower than those for life or CI insurance with underwriting based on both genetic test results and family history. Under the same circumstances, IPI applicants are more likely to be covered than those seeking life and CI insurance.

(b) If underwriting is based on genetic test, most mutation BRCA1 carriers are insurable at increased rates with DPs 1, 4 or 13 weeks. However, with DP 26 weeks, these persons become insurable for low ages and short policy terms. For mutation BRCA2 carriers, all premium increases are below 50%, well within the typical limit for insurability of 250% of the standard rates. Therefore, all cases for BRCA2 mutation carriers would be insurable. Considering ascertainment bias, we reduce the detection rates of OC by 50%. This leads to reduction of IPI premium rates for both mutation BRCA1/2 carriers as expected. Most cases, including mutation BRCA1 carriers seeking an IPI policy with DP 26 weeks, are insurable but at various levels of increased rates.

(c) Under family history underwriting, because of the 'averaging-out' effect, the premium rates offered for persons with a family history are lower than those for known mutation carriers. Most cases with a family history are insurable. Plausible reduction of detection rates of OC will allow more persons to be covered as standard risk.

(d) Under a moratorium on genetic test results and family history, the costs of adverse selection are all very small and negligible. Even in a small market with severe

adverse selection and exceptionally high genetic testing rate (0.035 per annum), the increase in premium is less than or slightly higher than 0.01%.

8.2 Further Research

We propose some suggestions for further research:

(a) Lu, Macdonald & Waters (2008) and Lu *et al.*, (2008) studied the impact of genetic information relevant to BC on the IPI market. Our work presented in Chapter 6 and 7 studied the similar topic relevant to OC. Following the methodology of the "bottom-up" approach demonstrated in Chapter 4, in order to quantify the overall impact on the IPI market, we also need to carry out more individual studies on other genetic disorders of interest, e.g. HNPCC and FAP, before we proceed to the overall impact of genetic information on the IPI market.

(b) In modeling insurance applicants' purchasing behaviours in the insurance market, we assume that in the case of moderate adverse selection, the purchase rate is doubled and in the case of severe adverse selection, the purchase rate is 0.25 per annum. These extreme assumption might be unrealistic, because consumers' behaviours are largely unknown, assuming that behaviour is governed by utility theory. Macdonald & Tapadar (2010) is good start, but more research work needs to be carried in this area.

(c) The causes of multifactorial disorders, e.g. heart disease and cancers, depend on the interaction between genetics and environment. However, this interaction is still not very clear to us. It will be very interesting to explore this area.

(d) Longevity risk has been an urgent issue for long-term annuity business. Apart from looking for statistical evidence to analyse the trend of future mortality improvement,

the cause of death and longevity were explored from different perspectives, of which genetics is one of them. The mechanism linking genetics and longevity is still largely unknown and we would be interesting to see the interaction between genetics and longevity.

(e) Modelling the impact of adverse selection in a economic framework has been discussed in Macdonald & Tapadar (2010) and Macdonald & McIvor (2009), especially the speculative adverse selection spiral, which might lead to the market collapse. Although we did not see any evidence of an adverse selection spiral in practice in response to the increasing knowledge of genetics, we shall prove the impossibility of adverse selection spiral in theory.

(f) In a broad sense, this thesis provides statistical evidence for underwriting insurance applicants at risk of genetic disorders. Our work coincides with the direction of 'evidence-based underwriting'. We have already seen that Sex Discrimination has taken effect, which governs the insurance underwriting using gender as a risk factor. Currently, the insurance industry has used many other risk factors in underwriting practice, such as, age, smoking status, disability, life style, postcode, etc. Strong evidences support the differentiation of market by gender and age. However, other risk factors, e.g. life style and postcode, are less clear-cut. In addition, in consideration of multifactorial genetic disorders, more statistical work should be carried out in these areas in support of underwriting in practice.

Appendix A

Intensity of Contracting Other Critical Illness and Adjusted Intensity of Death

The estimation of the intensities μ_x^{02} and μ_x^{03} has been provided by Gutiérrez & Macdonald (2003). These intensities are used in Chapters 2, 3 and 4.

- Estimation of μ_x^{02}

  From (1), μ_x^{02} is the sum of four intensities, μ_x^{02h}, μ_x^{02c}, μ_x^{02s} and μ_x^{02o} (Gutiérrez & Macdonald, 2003).

  1. For males:

$$\begin{aligned} \mu_x^{02c} &= \exp(-11.25 + 0.105x)(x < 51), \\ \mu_x^{02c} &= 0.2591585 - 0.01247354x + 0.0001916916x^2 \\ &\quad -8.952933 \times 10^{-7} x^3 (x \geqslant 60), \end{aligned}$$

     with a blending by linear interpolation between ages 51 and 60.

     By linear interpolation, we mean interpolation between both functions at each

age, for example

$$\mu_{55}^{02c} = \frac{[(60-55)\mu_{55}^{1} + (55-51)\mu_{55}^{2}]}{60-51}.$$

For females:

$$\begin{aligned} \mu_x^{02c} &= \exp(-10.78 + 0.123x - 0.00033x^2)(x < 53), \\ \mu_x^{02c} &= \exp(-0.01545632 + 0.0003805097x)(x \geqslant 53). \end{aligned}$$

2. For males:

$$\begin{aligned} \mu_x^{02h} &= \exp(-13.2238 + 0.152568x)(x < 44), \\ \mu_x^{02h} &= \exp(-0.01245109 + 0.000315605x)(x > 49), \end{aligned}$$

with a blending by linear interpolation between ages 44 and 49.

For females:

$$\mu_x^{02h} = \frac{0.598694}{\Gamma(15.6412)} \times 0.15317^{15.6412} \exp(-0.15317)x^{14.6412}.$$

In practice, the onset of a heart attack will not result in an instant CI benefit payment. The patient has to survive for another 28 days after onset to get the payment. Let p_x^h be the 28-day survival probability after the first-ever heart attack. We take 28-day mortality rates following a heart attack ($q_x^h = 1 - p_x^h$) from Dinani *et al.* (2000). For females $q_x^h = 0.21$ at ages 20 – 80. The rates for males are given in Table 1.

3. For males:

$$\mu_x^{02s} = \exp(-16.9524 + 0.294973x - 0.001904x^3 + 0.00000159449x^3).$$

Table A.1: 28-day mortality rates ($q_x^h = 1 - p_x^h$) for males following heart attack.

Age	q_x^h	Age	q_x^h	Age	q_x^h	Age	q_x^h
20–39	0.15	47–52	0.18	58–59	0.21	65–74	0.24
40–42	0.16	53–56	0.19	60–61	0.22	75–79	0.25
43–46	0.17	57	0.20	62–64	0.23	80+	0.26

For females:

$$\mu_x^{02s} = \exp(-11.1477 + 0.08107x).$$

Following a stroke, the 28-day survival probabilities p_x^s, are taken from Dinani *et al.* (2000). For males and females $p_x^s = (0.9 - 0.002x)/0.9$.

4. Following Macdonald, Waters & Wekwete (2003b) and Dinani *et al.* (2000), we suppose that other minor causes of CI insurance claims amount to 15% of those arising from cancer, heart attack, and stroke. Therefore:

$$\mu_x^{02o} = 0.15(\mu_x^{02c} + p_x^h \times \mu_x^{02h} + p_x^s \times \mu_x^{02s}).$$

By summing all four intensities, we have:

$$\mu_x^{02} = 1.15(\mu_x^{02c} + p_x^h \times \mu_x^{02h} + p_x^s \times \mu_x^{02s}).$$

- Estimation of μ_x^{03}

  Mortality μ_x^{03} is based on the English Life Tables No.15 (μ_x^{ELT15}) with mortality from causes leading to CI claims removed. We added back the 28-days mortality following heart attack and strokes as follows:

$$\mu_x^{03} = (1 - \theta_x)\mu_x^{ELT15} + (1 - p_x^h)\mu_x^{02h} + (1 - p_x^s)\mu_x^{02s},$$

where for males:

$$\begin{aligned}\theta_x &= 0.0185408 + 0.0655723x - 0.00667150x^2 \\ &\quad +0.000223974x^3 - 0.00000228356x^4 (x < 30), \\ \theta_x &= -2.80056 + 0.149759x \\ &\quad -0.00203616x^2 + 0.00000881081x^3 (x > 44),\end{aligned}$$

with a linear blending between ages 30 and 44,

For females:

$$\begin{aligned}\theta_x &= -0.0261291 + 0.104641x - 0.0118145x^2 \\ &\quad +0.000467135x^3 - 0.00000579010x^4 (x < 30), \\ \theta_x &= -1.34514 + 0.0897216x - 0.00119978x^2 \\ &\quad 0.00000486785x^3 (x > 35),\end{aligned}$$

with a linear blending between ages 30 and 35.

Appendix B

An Expanded CI Insurance Market Model of MD

Figure 2.6 shows the C.I. insurance market model adopted from Gutiérrez & Macdonald (2004). This model is deceptively simple in the case of MD. A more detailed expanded model for computation is shown in Figure B.1:

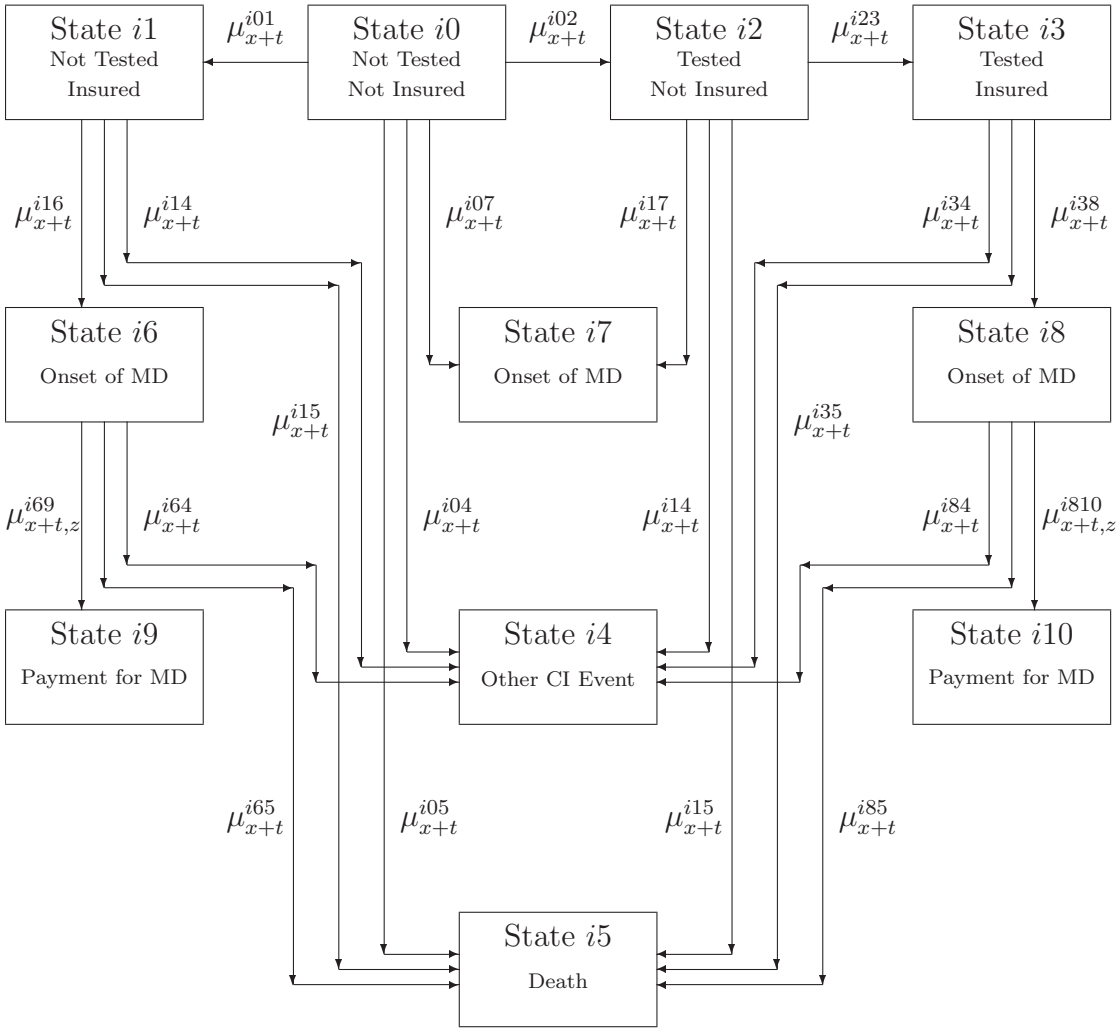

Figure B.1: An expanded semi-Markov model of insurance purchase and critical illness insurance events for a person in sub-population i.

Appendix C

Technical Details of Epidemiologies of Genetic Disorders

The Appendix C presents the technical details of genetic disorders we reviewed in Chapter 3.

C.1 Adult Polycystic Kidney Disease (APKD)

The Section C.1 correspondents with Section 3.1 in Chapter 3.

C.1.1 Onset Rates of ESRD for APKD1 and APKD2 Mutation Carriers

We use $q(x)$ to denote the penetrance at age x. As mentioned before, three studies have published penetrance estimates of APKD1 and APKD2 mutations: Johnson & Gabow (1997), Hateboer *et al.* (1999) and Ravine *et al.* (1992). These papers gave graphs of Kaplan-Meier estimates of the 'survival' probability $1 - q(x)$. In these cases, the event of interest was the first time to occur of ESRD, or death by any cause. That is, death

was not treated as a type of censoring. The approach we adopt is to fit suitable curves to each Kaplan-Meier estimate, which we take to be estimates of:

$$S(x) = \exp(-\int_0^x (\mu_t^{ESRD} + \mu_t^{DEAD})dt),$$

where μ_x^{ESRD} is the onset rate of ESRD and μ_x^{DEAD} is the force of mortality. Then μ_x^{ESRD} on its own can be found using a suitable population mortality table. In Gutiérrez & Macdonald (2007), English Life Table No. 15 is used to adjust. The fitted survival functions based on time to ESRD or death in Hateboer *et al.* (1999) is:

For APKD1:

$$S(x) = 1 - \exp(-9.08371 + 0.231087x - 0.00138536x^2)(x < 40)$$

$$S(x) = 23.0056 \left(\frac{0.345615^{15.1344} \exp(-0.345615x)x^{14.1344}}{\Gamma(15.1344)} \right) (x > 55),$$

with blending by sine curve between ages 40 and 55, and for APKD2:

$$S(x) = 1 - \exp(-11.8117 + 0.25559x - 0.00136435x^2)(x < 58)$$

$$S(x) = 24.7781 \left(\frac{0.364067^{21.278} \exp(-0.364067x)x^{20.278}}{\Gamma(21.278)} \right) (x > 70),$$

with blending by sine curve between ages 58 and 70. The numerical results are plotted in Figure 3.1.

C.1.2 Post-ESRD Mortality

The following are mathematical expressions of the Gamma functions fitted to Kaplan-Meier estimates of survival probability functions for ESRD patients receiving treatment, either dialysis or transplantation, based on which the intensity $\mu_{x,z}^{APKD,Mortality}$ is derived.

The fitted cumulative survival probability is

$$1 - S(z) = \frac{a^b}{\Gamma(b)} \int_0^z t^{b-1} e^{-at} dt,$$

with different choices of the parameters based on different age ranges. Variable z denotes the length of time patients have been in ESRD. The numerical results are plotted in Figures 3.2 and 3.3.

Table C.1: Parameterizations of Gamma function fitted to cumulative probability of duration of death following onset of dialysis or kidney transplant.

Ages at Onset	State	a	b
20-44	Dialysis	0.0442787	1.36159
20-44	Transplant	0.0746419	2.54771
45-59	Dialysis	0.0902674	1.61700
45-59	Transplant	0.0249060	1.26870

C.2 Early-onset Alzheimer's Disease (EOAD)

The Section C.2 correspondents with Section 3.2 in Chapter 3.

C.2.1 Onset Rates of EOAD for PSEN-1 Mutation Carriers

Espinosa-Castañeda (2006) produced empirical results for the onset rate of EOAD, which does not have a mathematical expression. Please refer to Espinosa-Castañeda (2006) for more details.

C.2.2 Post-onset Mortality Rate for PSEN-1 Mutation Carriers

The mathematical expression of post-onset mortality rate $\mu_z^{EOAD,Mortality}$ for PSEN-1 mutation carriers is:

$$\mu(z) = \left[0.012250264 z^{1.37601} \exp\left(-0.00168128 z^{2.37601}\right)\right],$$

where variable z denotes the length of time people have been in the EOAD state. The numerical results are plotted in Figure 3.7.

C.3 Huntington's Disease (HD)

The Section C.3 correspondents with Section 3.3 in Chapter 3.

C.3.1 Onset Rates of HD

Since we assume the age of onset, X to have a $N(\mu, \sigma)$ distribution, its density function is:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Denote its cumulative distribution function $F_X(x)$. So the onset rate is:

$$\begin{aligned} \mu_x(\mu, \sigma) &= \frac{f_X(x)}{1 - F_X(x)} \\ &= \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}}{1 - \Phi(\frac{x-\mu}{\sigma})}, \end{aligned}$$

where $\Phi(x)$ is the C.D.F of the unit normal distribution. In our case the estimated parameters are $\mu = 45.038543$ and $\sigma = 14.516176$ (MacCalman, 2009). We are mainly interested in ages between 20 and 60. The numerical results are plotted in Figure 3.8.

C.3.2 Post-Onset Mortality of HD Mutation Carriers

In Gutiérrez & Macdonald (2004), the probability of surviving for z years since onset of HD, $S(z)$, is defined as follows.

For age at onset 20–34:

$$1 - S(z) = \frac{0.174219^{4.11789}}{\Gamma(4.11789)} \int_0^z t^{3.11789} e^{-0.174219t} dt.$$

For age at onset 35–49:

$$1 - S(z) = \frac{0.177225^{4.35064}}{\Gamma(4.35064)} \int_0^z t^{3.35046} e^{-0.177225t} dt.$$

For age at onset 50 and over:

$$1 - S(z) = \frac{0.183372^{4.1465}}{\Gamma(4.1465)} \int_0^z t^{3.1465} e^{-0.183372t} dt.$$

The numerical results are plotted in Figure 3.9.

C.4 HNPCC

The Section C.4 correspondents with Section 3.4 in Chapter 3.

C.4.1 Estimation of Onset Rates in HNPCC

The following are the fitted cumulative onset probabilities of contracting CRC for mutation MLH1 and MSH2 carriers, males and females (Lu *et al.*, 2007).

$$F(x)^{CRC}_{MLH1.m} = 0.3406 - 0.039040x + 0.001294x^2 - 0.000009611x^3 (22 \leq x \leq 70),$$

in which we assume $F(x)^{CRC}_{MLH1.m} = 0$ below age 22.

$$F(x)^{CRC}_{MLH1.f} = 0.2193 - 0.02570x + 0.0008584x^2 - 0.000006075x^3 (23 \leq x \leq 70),$$

in which we assume $F(x)^{CRC}_{MLH1.f} = 0$ below age 23.

$$F(x)^{CRC}_{MSH2.m} = \frac{\exp\left(-10.92 + 0.3512x - 0.002579x^2\right)}{1 + \exp\left(-10.92 + 0.3512x - 0.002579x^2\right)} (20 \leq x \leq 68).$$

$$F(x)^{CRC}_{MSH2.f} = \frac{\exp\left(-12.36 + 0.3498x - 0.002421x^2\right)}{1 + \exp\left(-12.36 + 0.3498x - 0.002421x^2\right)} (20 \leq x \leq 70).$$

The following are the fitted cumulative onset probabilities of contracting EC for mutation MLH1 and MSH2 carriers, females only (Lu *et al.*, 2007).

$$F(x)^{EC}_{MLH1} = \frac{\exp(-17.78 + 0.4975x^1 - 0.003655x^2)}{1 + \exp(-17.78 + 0.4975x^1 - 0.003655x^2)} (20 \leq x \leq 65).$$

$$F(x)^{EC}_{MSH2} = \frac{\exp(-4.307 - 0.1973x + 0.008763x^2 - 7.432 \times 10^{-5}x^3)}{1 + \exp(-4.307 - 0.1973x + 0.008763x^2 - 7.432 \times 10^{-5}x^3)} (20 \leq x \leq 65).$$

The following are the fitted cumulative onset probabilities of contracting OECC for mutation MLH1 and MSH2 carriers, males and females. OECC stands for extracolonic cancers, including the cancers of stomach, urinary tract, small bowel, ovary (female only) and brain (Lu *et al.*, 2007).

$$F(x)^{OECC}_{MLH1.m} = \frac{\exp\left(-30.6539 + 0.8128x^1 - 0.0058x^2\right)}{1 + \exp\left(-30.6539 + 0.8128x^1 - 0.0058x^2\right)}(20 \leq x \leq 70).$$

$$F(x)^{OECC}_{MLH1.f} = \frac{\exp\left(-18.76 + 0.4747x - 0.003334x^2\right)}{1 + \exp\left(-18.76 + 0.4747x - 0.003334x^2\right)}(20 \leq x \leq 70).$$

$$F(x)^{OECC}_{MSH2.m} = \frac{\exp\left(-7.1635 + 0.10027x - 2.3206 \times 10^{-4}x^2\right)}{1 + \exp\left(-7.1635 + 0.10027x - 2.3206 \times 10^{-4}x^2\right)}(42 \leq x \leq 65),$$

in which below age 42, the intensity corresponding to $F(x)^{OECC}_{MSH2.m}$ is extrapolated linearly to the origin.

$$F(x)^{OECC}_{MSH2.f} = \frac{\exp\left(-10.45 + 0.2501x - 0.001618x^2\right)}{1 + \exp\left(-10.45 + 0.2501x - 0.001618x^2\right)}(20 \leq x \leq 70).$$

The following are the fitted onset rates of CRC for populations, both males and females (Lu *et al.*, 2007).

$$\mu^{CRC}_{pop.m}(x) = 0.001401\frac{\Gamma(10.196)}{\Gamma(8.196)\Gamma(x)}\left(\frac{80x}{89^2}\right)^{7.196}\left(1 - \frac{80x}{89^2}\right)$$

$$\mu^{CRC}_{pop.f}(x) = 0.001092\frac{\Gamma(8.207)}{\Gamma(6.742)\Gamma(1.465)}\left(\frac{80x}{89^2}\right)^{5.742}\left(1 - \frac{80x}{89^2}\right)^{0.465}$$

The following are the fitted onset rates of EC for populations, females only (Lu *et al.*, 2007).

$$\mu_{pop}^{EC}(x) = \begin{cases} \exp(-17.32 - 0.09261x + 0.004273x^2 - 5.200 \times 10^{-5}x^3) & (20 \leq x \leq 54) \\ -4.665 \times 10^{-4} + 2.446 \times 10^{-5}x - 1.555 \times 10^{-7}x^2 & (57 \leq x \leq 89), \end{cases}$$

blended linearly between age 54 and 57.

C.4.2 Estimation of Post-onset Mortality in HNPCC

We surveyed the following papers, in order to estimate the post-onset mortality associated with CRC.

1. Watson *et al.* (1998)

   In this case-control study, the study group was the sampled HNPCC patients and the control group was the sporadic CRC patients. In the study group, CRC cases were selected from 98 HNPCC families in the registries at Roswell Park Cancer Institute and Creighton University. Amsterdam criteria is used to ascertain the HNPCC patients. In the control group, the sporadic CRC patients were selected from the tumor registry (TR) of a single hospital affiliated with Creighton University. Every patient was staged at the time of diagnosis based on TNM system. The diagnosis of CRC was treated as the start of survival analysis. The observation ends either by censoring because of death, or development of CRC, or 10 years after the analysis. The Kaplan-Meier methods were used in the survival analysis. We obtained the original data from Dr. Patrice Watson, the principal author of Watson *et al.* (1998).

2. Sankila *et al.* (1996)

   Sankila *et al.* (1996) is a case-control study, where the sampled HNPCC patients represent the study group and sampled sporadic CRC patients represent the study group. In the study group, 175 HNPCC patients from 39 families were ascertained

by using Amsterdam Criteria. Genetic testing was carried out in these families. One similar germline mutation of MLH1 has been found in 17 of these families, another mutation has been found in 4 families. Of the 175 patients, 120 were from the families with germline mutation in MLH1 gene. From 1953 to 1993, 14,261 sporadic CRC patients were reported to Finnish Cancer Registry. Of These people, 14,086 patients younger than 65 year old represent the control group. The follow-up will stop either at the date of death or emigration or on the closing day of December 31, 1993, whichever occurred first. No patient was lost from the follow-up. Cumulative relative survival rates (RSRs) were calculated by dividing the observed with the expected survival rates. The expected survival rates were derived from the sex, age and calendar period-specific life tables of the general Finnish population. The main conclusion drawn from this paper is that survival rate of HNPCC patients is better than sporadic CRC patients.

3. Lin *et al.* (1998)

   Lin *et al.* (1998) is a case-control study. The study group was composed of members from seven MLH1 germline mutation families and five MSH2 germline mutation families diagnosed with colorectal cancer from 1945 to 1991 and registered at the Hereditary Cancer Institute (Creighton University). Colorectal cancer patients from 1965 to 1996, totalling 1,189, registered with the Creighton University tumor registry, served as the general population cohort. Ten-year survival was calculated using Kaplan-Meier methodology. The result is that combined MLH1 and MSH2 ten-year survival was 68.7% compared with 47.8% for the general population. Therefore HNPCC patients indeed have a better survival probability than the sporadic CRC patients.

4. Tomoda, Baba & Akazawa (1999)

The authors compared the survival between 46 HNPCC patients (study group) and 1185 sporadic CRC patients registered at the National Kyushu Cancer Center between 1972 and 1995 (control group). Five-year survival probabilities were 78.1% and 62.2% for study group and control group respectively. The survival probabilities were calculated using Kaplan-Meier methods. The main conclusion is that the prognosis of HNPCC patients is better than that of sporadic CRC patients, and it agreed with the conclusion drawn in Sankila *et al.* (1996).

5. Bertario *et al.* (1999)

   The authors examined 2,340 colorectal-cancer patients: 144 HNPCC patients (Amsterdam Criteria), 161 FAP patients and 2,035 patients with sporadic cancer. Data on hereditary-cancer patients treated between 1980 and 1995 was collected in a registry. The 2,035 sporadic colorectal-cancer patients (controls) included all new cases treated in the Department of Gastrointestinal-Tract Surgery during the same period. Observed survival was estimated using the Kaplan-Meier method. Cumulative survival probability was estimated at 5 years within each group. All patients were staged using the Dukes system. In the sporadic group, 51% were early-stage cancers (Dukes A or B) vs. 48.4% and 52.1% in the FAP and HNPCC groups respectively. In the HNPCC, FAP and sporadic-cancer groups, the 5-year cumulative survival rates was 56.9%, 54.4% and 50.6% respectively. The survival rates were calculated using the Kaplan-Meier methods. The conclusion is that FAP patients, especially HNPCC patients appear to have a better prognosis than sporadic CRC patients.

6. Elsakov & Kurtinaitis (2006)

   The authors aimed to evaluate survival rates in Lithuanian HNPCC patients and compare them with survival rates of sporadic cases arising from the general population. This is a case-control study. The study group consisted of 8 patients from

6 HNPCC families, diagnosed between 1995 and 1999. HNPCC patients characteristic (age and stage) were used to trace the records of the Cancer Registry at the same period to identify the control cases corresponding the required criteria. 263 sporadic CRC patients were found - 106 at stage II and 157 at stage III. The result is that the 10-year survival was 87.5% in the HNPCC study group compared with only 44.8% in the control group. The 10-year survival probabilities were calculated using Kaplan-Meier methodology. The conclusion is that HNPCC patients with confirmed MSH2 or MLH1 mutations diagnosed with stages II and III CRC have a good 10-year survival prognoses compared with those from the general population.

7. Percesepe *et al.* (1997)

   This study evaluated the clinical outcome of HNPCC patients with respect to that of patients with colorectal cancer recorded in a population-based cancer registry. The authors assessed survival of 85 colorectal cancer patients from 24 unrelated families defined as having HNPCC according to the criteria of the International Collaborative Group and a 5-year follow-up (cancer diagnosis from 1980-1989) were available. 377 sporadic CRC patients, registered from 1984-1986, with a 5-year follow-up, were used for comparison. Overall, Colorectal cancer-specific 5-year survival rates were 55.2% and 42.5% for HNPCC patients and sporadic CRC patients, respectively. The patients were stage-stratified and compared in terms of 5-year survival probabilities. Stage II HNPCC patients exhibited a better prognosis than the corresponding cases in the sporadic CRC patients. Other stages were incomparable due to the small number of HNPCC patients. The survival rates were calculated using Kaplan-Meier methods. The conclusion is both overall and stage II HNPCC patients showed a survival advantage in comparison with sporadic CRC patients.

We surveyed a series of papers, in order to estimate the post-onset mortality associated with OECC, including cancer of stomach, urinary tract, small bowel, ovarian (female

only) and brain. However, only three papers were located, which related to stomach cancer, small bowel cancer and ovarian cancer. As to HNPCC-associated urinary tract cancer and brain cancer, we can not locate any paper.

1. Ovarian Cancer: Crijnen *et al.* (2005)

   The aim of this study was to compare the survival after contracting OC between HNPCC patients (study group) and sporadic OC patients (control group). A total of 26 HNPCC patients with OC, as study group, were identified from the Dutch HNPCC Registry. A control group (52 cases) matched for age, stage and year of diagnosis was derived from the population-based Eindhoven Cancer Registry. Kaplan-Meier methods were used to calculate the survival probabilities. The authors found the cumulative 5-year-survival rates were 64.2 and 58.1% respectively, and concluded that the survival rates was not significantly different between HNPCC patients with OC and the sporadic OC patients.

2. Gastric Cancer: Aarnio *et al.* (1997)

   The authors gathered clinical data relating to patients recorded in the Finnish HNPCC registry, in order to identify characteristics of HNPCC-associated gastric cancer. This study is not a case-control study. The data included 51 families with a characterized mutation and/or that met the Amsterdam criteria. Of 570 members affected by malignancy, gastric cancer occurred in 62. The overall 5-year survival rate was 15%. The 5-year survival rate was 48% in cases in whom radical surgery had been undertaken. Kaplan-Meier methods were used to calculate the survival rates. However, this paper did not include the information about the 5-year survival rate for the sporadic gastric cancer.

3. Small bowel cancer: Schulmann *et al.* (2005)

The authors aimed to study the risk of HNPCC-associate small bowel cancer (HNPCC-SBC). In the study, the information of 32 HNPCC-SBC patients were retrieved from the databased of the German HNPCC Consortium based on the Amsterdam and Bethesda criteria. The overall 10-year survival rate was 87%. Kaplan-Meier methods were used to calculate the survival rate. However, this paper did not include the information about the 5-year survival rate for the sporadic small bowel cancer.

C.5 BC & OC

The Section C.5 correspondents with Section 3.5 in Chapter 3.

C.5.1 Onset Rates of Breast Cancer and Ovarian Cancer

The following are the fitted onset rates of breast cancer and ovarian cancer for mutation BRCA1 and BRCA2 carriers (Gui *et al.*, 2006).

$$\mu_x^{BC,BRCA1} = 1.1874 \times 10^{-16} e^{-0.21x} x^1 1.2,$$

$$\mu_x^{BC,BRCA2} = 8.3108 \times 10^{-13} e^{-0.1x} x^7 .37,$$

$$\mu_x^{OC,BRCA1} = 1.3318 \times 10^{-9} e^{-0.03x} x^4 .48,$$

$$\mu_x^{OC,BRCA2} = 3.5915 \times 10^{-79} e^{1.00x} x^5 6.95.$$

The following are the fitted onset rates of breast cancer and ovarian cancer for non-mutation carriers (Gui *et al.*, 2006).

$$\mu_x^{BC,POP} = \begin{cases} 6.0425 \times 10^{-15} e^{-0.0742} x^{7.7305} & (0 \le x \le 53) \\ 0.00012 + 0.00018(x-35) - 0.00005(x-35)^2 + 0.0000000529(x-35)^3 & (x \ge 53), \end{cases}$$

$$\mu_x^{OC,POP} = \begin{cases} 1.3567 \times 10^{-13} e^{-0.035x} x^5 .92 & (0 \le x \le 45) \\ 0.0001554 + 0.000029(x-45) - 0.00000048(x-45)^2 & (x \ge 55). \end{cases} \tag{C.1}$$

C.5.2 Post-onset Mortality Associated with Breast Cancer and Ovarian Cancer

The following are the graduated post-onset mortality associated with breast cancer and ovarian cancer for each duration separately up to 6 years, and then to 6 years and over (Gui *et al.*, 2006).

$0 \le Duration \le 1$

$$\begin{aligned} \mu_{x,d}^{BC} &= 2.00266 - 0.1507811x + 0.004264272x^2 - 5.27552 \times 10^{-5} x^3 + 2.456224 \times 10^{-7} x^4, \\ \mu_{x,d}^{OC} &= \exp(-2.71394 + 0.023657x + 0.1960156 \times 10^{-3} x^2). \end{aligned}$$

$1 \le Duration \le 2$

$$\begin{aligned} \mu_{x,d}^{BC} &= \exp(0.6037712037 - 0.01500175x + 0.1111315 \times 10^{-3} x^2) \\ &\quad \times (0.0474102 + 0.307835 \times 10^{-3} x + \exp(3.06993 - 0.284105x + 0.00266558x^2)), \\ \mu_{x,d}^{OC} &= \exp(-9.6097147 + 0.3634146x - 0.005149204x^2 + 0.2471276 \times 10^{-4} x^3). \end{aligned}$$

$2 \le Duration \le 3$

$$\begin{aligned} \mu_{x,d}^{BC} &= \exp(0.4971436820 - 0.01500175x + 0.1111315 \times 10^{-3} x^2) \\ &\quad \times (0.0474102 + 0.307835 \times 10^{-3} x + \exp(3.06993 - 0.284105x + 0.00266558x^2)), \\ \mu_{x,d}^{OC} &= \exp(-10.1965014 + 0.3634146x - 0.005149204x^2 + 0.2471276 \times 10^{-4} x^3). \end{aligned}$$

$3 \leq Duration \leq 4$

$$\begin{aligned}
\mu_{x,d}^{BC} &= \exp(0.3905161603 - 0.01500175x + 0.1111315 \times 10^{-3}x^2) \\
&\quad \times (0.0474102 + 0.307835 \times 10^{-3}x + \exp(3.06993 - 0.284105x + 0.00266558x^2)), \\
\mu_{x,d}^{OC} &= \exp(-13.4719011 + 0.52647732x - 0.008227498x^2 + 0.4354431 \times 10^{-4}x^3).
\end{aligned}$$

$4 \leq Duration \leq 5$

$$\begin{aligned}
\mu_{x,d}^{BC} &= \exp(0.352482834 - 0.003144911x) \\
&\quad \times (0.02902753 + \exp(-0.1624326x + 0.00164027x^2)), \\
\mu_{x,d}^{OC} &= \exp(-14.1632748 + 0.52647732x - 0.008227498x^2 + 0.4354431 \times 10^{-4}x^3).
\end{aligned}$$

$5 \leq Duration \leq 6$

$$\begin{aligned}
\mu_{x,d}^{BC} &= \exp(0.082950962 - 0.000880887x) \\
&\quad \times (0.02902753 + \exp(-0.1624326x + 0.00164027x^2)), \\
\mu_{x,d}^{OC} &= \exp(-14.8546485 + 0.5384382x - 0.008227498x^2 + 0.4354431 \times 10^{-4}x^3).
\end{aligned}$$

$6 \leq Duration$

$$\begin{aligned}
\mu_{x,d}^{BC} &= \exp(0.082950962 - 0.000880887x) \\
&\quad \times (0.02902753 + \exp(-0.1624326x + 0.00164027x^2)), \\
\mu_{x,d}^{OC} &= \exp(-14.8546485 + 0.5384382x - 0.008227498x^2 + 0.4354431 \times 10^{-4}x^3).
\end{aligned}$$

C.5.3 Mortality Rates Excluding the Death Caused by BC and OC

For females, the fitted function for r_x^{BCOC} is

$$r_x^{BCOC} = \begin{cases} 2.0785 \times 10^{-17} e^{-0.28} x^{13.05} & (0 \leq x \leq 54) \\ 1.30144 - 0.02850194x + 0.0001588314x^2 & (x \geq 65), \end{cases}$$

with linear blending of the two functions between ages 54 and 65 (Gui *et al.*, 2006).

Appendix D

Fourth Order Runge-Kutta Method

Fourth Order Runge-Kutta methods are one of the most commonly used methods to solve ODE. We give a short summary here. See Press *et al.* (1988) for details.

Let $y' = f(x, y)$ and the initial value is that $y(x_0) = y_0$. Suppose h is the stepsize, we have the following recursive equations:

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}h(k_1 + 2k_2 + 2k_3 + k_4) \\ x_{n+1} &= x_n + h, \end{aligned}$$

where y_{n+1} is the approximation of $y(x_{n+1})$, and

$$\begin{aligned} k_1 &= f(x_n, y_n) \\ k_2 &= f(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1) \\ k_3 &= f(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2) \\ k_4 &= f(x_n + h, y_n + hk_3) \end{aligned}$$

The analogy is that the next value y_{n+1} is determined by the previous value y_n plus

the increment between x_n and x_{n+1}, which is the product of stepsize h and an estimated slope $\frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$.

References

Aaltonen, L., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R., Kaarianen, H., Eskelinen, M., Jarvinen, H., Mecklin, J. and Chapelle, A. (1998). Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *The New England Journal of Medicine*, **338**, 14811487.

Aarnio, M., Salovaara, R., Aaltonen, L.A., Mecklin, J.P. and Järvinen, H.J. (1997). Features of gastric cancer in hereditary non-polyposis colorectal cancer syndrome. *International Journal of Cancer*, **74**, 551–555.

Abeliovich, D., Lerer, I., Pashut-Lavon, I., Shmueli, E., Raas-Rothschild, A. and Frydman, M. (1993). Negative expansion of the myotonic dystrophy unstable sequence. *American Journal of Medical Genetics*, **52**, 1175–1181.

Adami H.O., Bergstrom, R., Persson, I., and Sparen, P. (1990). The incidence of ovarian cancer in Sweden, 1960 – 1984. *Am J Epidemiol*, **132**, 446–452.

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

Amos, C.I. and Struewing, J.P. (1993). Genetic epidemiology of epithelial ovarian cancer. *Cancer*, **71**, 566.

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Ponder, B.J., and Easton, D.F. (2001). Evidence for further breast cancer susceptibility genes in

addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology*, **21**, 1–18.

Armstrong, J., Boada, M., Rey, M., and Vidal, N. and Ferrer, I. (2004). Familial alzheimer's disease associated with a713t mutation in app. *Neuroscience Letters*, **370**, 241–243.

Ashizawa, T. (2000). New nomenclature and DNA testing guidelines for myotonic dystrophy type 1. *Neurology*, **54**, 1218–1221.

Ashizawa, T., Dubel, J.R., Dunne, P.W., Dune, C.J., Fu, Y.H., Pizzuti, A., Caskey, C.T., Boerwinkle, E., Perryman, M.B., Epstein, H.F. and Hejtmancik, J.H. (1992). Anticipation in myotonic dystrophy. II. complex relationships between clinical findings and structure of the CTG repeat. *Neurology*, **42**, 1877–1883.

Ashizawa, T., Dunne, P.W., Ward, P.A., Seltzer, W.K. and Richards, C.S. (1994). Effects of the sex of myotonic dystrophy patients on the unstable triplet repeat in their affected offspring.. *Neurology*, **44**, 120–122.

Astler, V.B. and Coller, F.A. (1954). The prognostic significance of direct extension of carcinoma of the colon and rectum. *Am Surg*, **139**, 846–852.

Auranen A. and Iselius, L. (1998). Segregation analysis of epithelial ovarian cancer in Finland. *Ann Hum Genet*, **77**, 1537–1541.

Banks, E., Beral, V. and Reeves, G. (1997). The epidemiology of epithelial ovarian cancer: A review. *International Journal of Gynecological Cancer*, **7**, 425–438.

Bell, J. (1934). Huntington's Chorea. *Treasury of Human Inheritance*, **IV**, 1–67.

Bell, J. (1947), Dystrophia myotonic and allied diseases, in *Treasury of Human Inheritance 4, Part V*, ed. Cambridge University Press, Cambridge

Bell, J., Brady, M.F., Young, R.C., Lage, J., Walker, J.L., Look, K.Y., Rose, G.S. and Spirtos, N.M. (2006). Randomized phase III trial of three versus

six cycles of adjuvant carboplain and paclitaxel in early stage epithelial ovarian carcinoma: A genecologic oncology group study. *Gynecologic Oncology*, **102**, 432–439.

Bellacosa, A., Genuari, M. and Anti, M. (1996). Hereditary nonpolyposis colorectal cancer: Review of clinical, molecular genetics, and counseling aspects. *American Journal of Medical Genetics*, **62**, 353-364.

Benedet, J.L. (2000). *Staging classifications and clinical practice guidelines for gynaecologic cancers*. Elsevier.

Beral, V., Fraser, P. and Chilvers, C. (1978). Does pregnancy protect against ovarian cancer?. *Lancet*, **20**, 1083–1097.

Bertario, L., Russo, A., Sala, P., Eboli, M., Radice, P., Presciuttini, S., Andreola, S., Rodriguez-Bigas, M.A., Pizzetti, P., Spinelli, P. (1999). Survival of patients with hereditary colorectal cancer: comparison of HNPCC and colorectal cancer in FAP patients with sporadic colorectal cancer. *International Journal of Cancer*, **80**, 183–187.

Bird, D. (2005). Alzheimer disease overview. Technical report. *University of Washington*.

Bliss, C.M. & Schroy, P.C. (2004). Endoscopic diagnosis and management of hereditary nonpolyposis colorectal cancer. *Current Opinion in Gastroenterology*, **20**, 468–473.

Boks, E.S., Trujillo, A.P., Voogd, A.C., Morreau, H., Kenter, G.G. and Vasen, H.F.A. (2002). Survival analysis of endometrial carcinoma associated with hereditary nonpolyposis colorectal cancer. *International Journal of Cancer*, **102**, 198–200.

Bouchard, G., Roy, R., Declos, M., Kouladjian, K. and Mathieu, J. (1988). Spreading of the gene for myotonic dystrophy in Saguenay (Quebec). *Journal de Génétique Humaine*, **36**, 221–237.

Booth, M., Beral, V. and Smith, P. (1989). Risk factors for ovarian cancer: A case control study. *Br J Cancer*, **60**, 592–598.

Brackenridge, R. and Elder, J. (1998). *Medical Selection of Life Risks*. 4th ed. New York: Macmillan.

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

Brunner, H.G., Nillesen, W., A van Oost, B., Jansen, G., Wieringa, B., Ropers, H.H. and Smeets, H.J.M. (1992). Presymptomatic diagnosis of myotonic dystrophy. *J Med Genet*, **29**, 780–784.

Buxton, J., Shelbourne, P., Davies, J., Jones, C., van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R. and Johnson, K. (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature*, **355**, 547–548.

Campion, D., Brice, A., Dumanchin, C., Puel, M., Baulac, M., De La Sayette, V., Hannequin, D., Duyckaerts, C., Michon, A., Martin, C., Moreau, V., Penet, C., Martinez, M., Clerget-Darpoux, F., Agid, Y., and Frebourg, T. (1996). A novel presenilin 1 mutation resulting in familial Alzheimer's disease with an onset age of 29 years. *NeuroReport*, **7**, 1582–1584.

Churchill, D., Bear, J., Morgan, J., Payne, R., McManamon, P. and Gault, H. (1984). Prognosis of adult onset polycystic kidney disease re-evaluated. *Kidney International*, **26**, 190–193.

Claus, E.B., Schildkraut, J.M., Thompson, W.D. and Risch, N.J. (1996). The genetic attributable risk of breast and ovarian cancer. *Cancer*, **77(11)**, 2318–2324.

Colombo, N., Chiari, S., Maggioni, A., Bocciolone, L., Torri, V. and Mangioni, C. (1994). Controversial issues in the management of early epithelial ovarian

cancer: Conservative surgery and role of adjuvant therapy. *Gynecologic Oncology*, **55**, S47–S51.

Continuous Mortality Investigation Committee (1991). Continuous Mortality Investigation Reports: Number 12. *The Institute of Actuaries and the Faculty of Actuaries*.

Continuous Mortality Investigation Committee (1993). Continuous Mortality Investigation Reports: Number 13. *The Institute of Actuaries and the Faculty of Actuaries*.

Continuous Mortality Investigation Committee (2001). Continuous Mortality Investigation Reports: Number 20. *The Institute of Actuaries and the Faculty of Actuaries*.

Covens, A., Carey, M., Bryson, P., Verma, S., Fung, M.F.K. and Johnston, M. (2002). Systematic review of first-line chemotherapy for newly diagnosed postoperative patients with stage II, III or IV epithelial ovarian cancer. *Gynecologic Oncology*, **85**, 71–80.

Crijnen, T.E., Janssen-Heijnen, M.L., Gelderblom, H., Morreau, J., Nooij, M.A., Kenter, G.G. and Vasen, H.F. (2005). Survival of patients with ovarian cancer due to a mismatch repair defect. *Familial cancer*, **4**, 301–305.

Dalgaard, O. (1957). Bilateral polycystic disease of the kidneys: A follow up of two hundred and eighty four patients and their families. *Acta Medica Scandinavica*, **S328**, 1–255.

Dinani, A., Grimshaw, D., Robjohns, N., Somerville, S., Spry, A. and Staffurth, J. (2000). A critical review: Report of the critical illness healthcare study group. *London: Staple Inn Actuarial Society*.

de Die-Smulders, C.E.M., Höweler, C.J., Thijs, C., Mirandolle, J.F., Anten, H.B., Smeets, H.J., Chandler, K.E. and Geraedts, J.P. (1998). Age

and causes of death in adult-onset myotonic dystrophy. *Brain*, **121**, 1557–1563.

de Jong, P. and Ferris, S. (2006). Adverse selection spirals. *Astin Bulletin*, **36(2)**, 589–628.

Department of Trade and Industry (1998). Government response to the Human Genetic Advisory Commission's Report on the implications of genetic testing for insurance. *DTI*.

DeVita, V.T. Jr., Hellman, S. and Rosenberg, S.A. (1997). *Cancer: principles and practice of oncology, 5th ed.* Philadelphia: Kippincott - Raven.

de Vos tot Nederveen Cappel, W.H., Nagengast, F.M., Griffioen, G., Menko, F.H., Taal, B.G., Kleibeuker, J.H. and Vasen, H.F. (2002). Surveillance for hereditary nonpolyposis colorectal cancer: a long-term study on 114 families. *Dis Colon Rectum*, **45**, 1588–1594.

Dizon, D.S., Weitzen, S., Rojan, A., Schwartz, J., Miller, J., Disilvestro, P., Gordinier, M.E., Moore, R., Tejada-Berges, T., Pires, L., Legare, R., Granai, C.O. (2006). Two for good measure: six versus eight cycles of carboplatin and paclitaxel as adjuvant treatment for epithelial ovarian cancer. *Gynecologic Oncology*, **100**, 417-421.

DiSaia, P.J. and Creasman, W.T. (2007). *Clinical Gynecologic Oncology*. Mosby.

Elsakov, P. and Kurtinaitis, J. (2006). Survival from Colorectal Carcinoma in HNPCC Families as Compared to the General Population in Lithuania – Initial Results. *Fam Cancer*, **5**, 369–371.

Espinosa-Castañeda, C. (2006). Ascertainment bias in estimating rates of onset of early-onset alzheimer's disease: a critical illness and life insurance application. PhD Thesis. *Heriot-Watt University, Edinburgh*.

Espinosa, C. and Macdonald, A.S. (2007). A correction for ascertainment bias in estimating rates of onset of highly penetrant genetic disorders. *ASTIN Bulletin*, **37**,

429–452.

Fader, A.N. & Rose, P.G. (2007). Role of surgery in ovarian carcinoma. *J Clin Oncol*, **25**, 2873–2883.

Farrington, S.M., McKinley, A.J., Carothers, A.D., Cunningham, C., Bubb, V.J., Sharp, L., Wyllie, A.H. and Dunlop, M.G. (1998). Evidence for an age-related influence of microsatellite instability on colorectal cancer survival. *International Journal of Cancer*, **98**, 844–850.

FIGO Committee on Gynecologic Oncology (2000). Staging classifications and clinical practice: Guidelines for gynaecologic cancers. `http://www.wvcancerregistry.org/Portals/27/PDFs/staging_booklet_figo.pdf`.

Ford, C., Kidd, A. and Hammond-Tooke, G. (2006). Myotonic dystrophy in Otago, New Zealand. *N Z Med J*, **119**, U2145.

Foroud, T., Gray, J., Ivashina, J. and Conneally, M. (1999). Differences in duration of Huntington's disease based on age at onset. *Journal of Neurology, Neurosurgery and Psychiatry*, **66**, 52–56.

Franceschi, S., Negri, E., Tzonou, A., Booth, M., La Vecchia, C., Parazzini, F., Beral, V., Boyle, P. and Trichopoulos, D. (1991). Pooled analysis of 3 European case-control studies of epithelial ovarian cancer: III. Oral contraceptive use. *Int J Cancer*, **49**, 61–65.

Frohock, A.M. (2003). The CTG Expansion Mutation In Myotonic Dystrophy: Genotype-Phenotype Comparisons and the Relevance to Insurance. Final Year Honours Project. *University of Nottingham, Nottingham.*

Gayther, S.A., Mangion, J., Russell, P., Seal, S., Barfoot, R., Ponder, B.A., Stratton, M.R. and Easton, D. (1997). Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. *Nat Genet*, **15**, 337–355.

Gebert, J., and Knebel Doeberitz, C.V. (1999). Hereditäre Tumoren des Gastrointestinaltrakts (FAP/HNPCC). *Chirurgische Gastroenterologie*, **15**, 163–172.

Giacomello, M., Barbiero, L., Zatti, G., Squitti, R., Binetti, G., Pozzan, T., Fasolato, C., and Guidoni, R. and Pizzo, P. (2005). Reduction of ca^{2+} stores and capacitive ca^{2+} entry is associated with familial alzheime's disease presenilin-2 t122r mutation and anticipates the onset of dementia. *Neurobiology of Disease*, **18**, 638–648.

Goldman, A., Ramsay, M. and Jenkins, T. (1994). Absence of myotonic dystrophy in southern African Negroids is associated with a significantly lower number of CTG trinucleotide repeats. *J Med Genet*, **31**, 37–40.

Grimm, T. (1975a). The age of onset and at death in dystrophia myotonica. *Journal de Génétique Humaine*, **23 SUPPL**, 172.

Grimm, T. (1975b). Myotonic dystrophy. *University of Göttingen*.

Gui, E.H. (2003). Modelling the impact of genetic testing on insurance – Early-onset Alzheimer's disease and other single-gene disorders. Ph.D Thesis. *Heriot-Watt University*.

Gui, E.H., Lu, B., Macdonald, A.S., Waters, H.R. and Wekwete, C. (2006). The genetics of breast and ovarian cancer III: A new model of family history with insurance applications. *Scandinavian Actuarial Journal*, **106**, 338–367.

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

Gutiérrez, C. and Macdonald, A.S. (2003). Adult Polycystic Kidney Disease and Critical Illness Insurance. *North American Actuarial Journal*, **7**, 93–115.

Gutiérrez, C. and Macdonald, A.S. (2004). Huntington's Disease, Critical Illness Insurance and Life Insurance. *Scandinavian Actuarial Journal*, **2004**, 279–313.

Gutiérrez, C. and Macdonald, A.S. (2007). Adult polycystic kidney disease and insurance: A case study in genetic heterogeneity. *North American Actuarial Journal*, **11**, 90–118.

Hamshere, M.G., Harley, H., Harper, P.S. Brook, J.D. and Brookfield, J.F.Y. (1999). Myotonic dystrophy: the correlation of (CTG) repeat length in leucocytes with age at onset in significant only for patients with small expansions. *Journal of Medical Genetics*, **36**, 59–61.

Hardcastle, J.D., Chamberlain, J.O., Robinson, M.H.E., Moss, S.M., Amar, S.S., Balfour, T.W., James, P. & Mangham, C.M. (1996). Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *The Lancet*, **348**, 1472–1477.

Harley, H.G., Rundle, S.A., MacMillan, J.C., Myring, J., Brook, J.D., Stephen, C., Reardon, W., Fenton, I. Shaw, D.J. and Harper, P.S. (1993). Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *American Journal of Human Genetics*, **52**, 1164–1174.

Harper, P.S. (1972). Genetic studies in myotonic dystrophy. PhD Thesis. *University of Oxford*.

Harper, P.S. (1973). Pre-symptomatic detection and genetic counselling in myotonic dystrophy. *Clin Genet*, **4**, 134–140.

Harper, P.S., Lim, C., and Craufurd, D. (2000). Ten years of presymptomatic testing for Huntington's disease: the experience of the UK Huntington's Disease Prediction Consortium. *Journal of Medical Genetics*, **37**, 657–571.

Harper, P.S. (1996). *Huntington's Disease*. WB Saunders, London.

Harper, P.S. (2001). *Myotonic Dystrophy*. London: WB Saunders.

Hashem, V.I. and Sinden R.R. (2002). Chemotherapeutically induced deletion of

expanded triplet repeats. *Mutation Research*, **508**, 107–119.

Hateboer, N., Bogdanova, N., Dijk, M., Coto, E., Saggar-Malik, A., San Millan, J., Torra, R., Ravie, D. & Breuning, M., for the European PKD1-PKD2 Study Group (1999). Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *The Lancet*, **353**, 103–107.

Herzog, T.J., Holloway, R.W. and Stuart G.C.E. (2003). Workshop: Options for therapy in ovarian cancer. *Gynecologic Oncology*, **90**, S45-S50.

Hewitson, P., Glasziou, P., Irwig, L., Towler, B. and Watson, E. (2007). Screening for colorectal cancer using the faecal occult blood test, Hemoccult (Cochrane Review). *Cochrane Database System Review*, **1**, .

Heyman, A., Wilkinson, W.E., Hurwitz, B.J., Helms, M.J., Haynes, C.S., Utley, C.M., and Gwyther, L.P. (1987). Early-onset Alzheimer's disease: clinical predictors of institutionalization and death. *Neurology*, **37**, 980–984.

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

Hoedema, R., Monroe, T., Bos, C., Palmer, S., Kim, D., Marvin, M. and Luchtefeld, M. (2003). Genetic testing for hereditary nonpolyposis colorectal cancer. *The American surgeon*, **69**, 387–391.

Högberg, T., Carstensen, J. and Simonsen, E. (1993). Treatment results and prognostic factors in a population-based study of epithelial ovarian cancer. *Gynecological Oncology*, **48**, 38-49.

Högberg, T., Glimelius, B., and Nygren, P. for the SBU-group (2001). A systematic overview of chemotherapy effects in ovarian cancer. *Acta Oncologica*, **40**, 340-360.

Höweler, C.J. (1986). A clinical and genetic study in myotonic dystrophy. PhD Thesis.

University of Rotterdam.

Höweler, C.J., Busch, H.F.M., Graedts, J.P.M., Niermeijer, M.F. and Staal, A. (1989). Anticipation in myotonic dystrophy: Fact or fiction. *1989*, **Brain**, 112.779–797

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

Hsiao, K.M., Chen, S.S., Li, S.Y., Chiang, S.Y., Lin, H.M., Pan, H., Huang, C.C., Kuo, H.C., Jou, S.B., Su, C.C., Ro, L.S., Liu, C.S., Lo, M.C., Chen, C.M. and Lin, C.C. (2003). Epidemiological and genetic studies of myotonic dystrophy type 1 in Taiwan. *Neuroepidemiology*, **22**, 283–289.

Hunter, A., Tsilfidis, C., Mettler, G., Jacob, P., Mahadevan, M., Surh, L. and Korneluk, R. (1992). The correlation of age of onset with CTG trimucleotide repeat amplification in myotonic dystrophy. *Journal of Medical Genetics*, **29**, 774–779.

Hunter, A.G.W., Jacob, P., O'Hoy, K., MacDonald, I., Mettler, G., Tsilfidis, C. and Korneluk, R.G. (1993). Decrease in the size of the myotonic dystrophy CTG repeat during transmission from parent to child: implications for genetic counselling and genetic anticipation. *American Journal of Medical Genetics*, **45**, 401–407.

Ikenaga, M., Tomita, N., Sekimoto, M., Ohue, M., Yamamoto, H., Miyake, Y., Mishima, H., Nishisho, I., Kikkawa, N. and Monden, M. (2002). Use of microsatellite analysis in young patients with colorectal cancer to identify those with hereditary nonpolyposis colorectal cancer. *Journal of Surgical Oncology*, **79**, 157–165.

International Collaborative Ovarian Neoplasm (ICON1) Collaborators (2003). International collaborative ovarian neoplasm trial 1: A randomized trial of adjuvant chemotherapy in women with early-stage ovarian cancer. *Journal of the*

National Cancer Institute, **95**, 125–132.

International Collaborative Ovarian eoplasm 1 (ICON1) and European Organization for Research and Treatment of Cancer Collaborators – Adjuvant Chemotherapy In Ovarian Neoplasm (EORTC–ACTION) (2003). International collaborative ovarian neoplasm trial 1 and adjuvant chemotherapy in ovarian neoplasm trial: Two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma. *Journal of the National Cancer Institute*, **95**, 105–112.

Järvinen H.J., Mecklin J.P., Sistonen P. (1995). Screening reduces colorectal cancer rate in families with hereditary. *Gastroenterology*, **108**, 1405–1411.

Järvinen, H.J., Aarnio, M., Mustonen, H., Aktan-Collan, K., Aaltonen, L.A., Peltomäki, P., De La Chapelle, A. and Mecklin, J.P. (2000). Controlled 15-year trial on screeing for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology*, **118**, 829–834.

Järvinen, H.J. (2006). Endoscopic surveillance in hereditary nonpolyposis colorectal cancer. *Techniques in Gastrointestinal Endoscopy*, **8**, 110–113.

Jacobs, I. and Lancaster, J. (1996). The molecular genetics of sporadic and familiar epithelian ovarian cancer. *Int J Gynecol Cancer*, **6**, 337–355.

Johnson, A. & Gabow, P. (1997). Identification of patients with autosomal dominant polycystic kidney disease at highest risk of end-stage renal disease.. *ournal of Medical Genetics*, **34**, 827–830.

Klein, D. (1958). La dystrophie myotonique (Steinert) et la myotonie congénitale (Thomsen) en Suisse: étude clinique, génétique et démographique. *Journal de Génétique Humaine*, **7 SUPPL**, 1–328.

Kristensen, G.B. and Tropé, C. (1997). Epithelian ovarian carcinoma. *Lancet*, **349**, 113–117.

Kronborg, O., Fenger, C. Olsen, J., Jørgensen, O.D. and Søndergaard, O. (1996). Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet*, **348**, 1467-1471.

Langlois, M. (2003). RNA-based gene therapies for myotonic dystrophy type 1. PhD Thesis. *Université Laval, Québec.*

Lin, K.M., Shashidharan, M., Thorson, A.G., Ternent, C.A., Blatchford, G.J., Christensen, M.A., Watson, P., Lemon, S.J., Franklin, B., Karr, B., Lynch, J. and Lynch, H.T. (1998). Cumulative incidence of colorectal and extracolonic cancers in MLH1 and MSH2 mutation carriers of hereditary nonpolyposis colorectal cancer. *J Gastrointest Surg*, **2**, 67–71.

Liu, B., Parsons, R., Papadpoulos, N. *et al.* (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nature Medicine*, **2**, 169–174.

López de Munain, A., Emparanza, J.I., Poza, J.J., Martí Massó, J.F., Cobo, A., Martorell, L., Baiget, M., and Martínez Lage, J.M. (1993). Prevalence of myotonic dystrophy in Guipúzcoa (Basque Country, Spain). *Neurology*, **43**, 1573–1576.

Lu, B.P., Macdonald, A.S. and Waters, H.R. (2008). The genetics of breast and ovarian cancer IV: A model of breast cancer progression. *Submitted.*

Lu, B.P., Macdonald, A.S., Waters, H.R. and Yu, F. (2008). The genetics of breast and ovarian cancer V: Application to income protection insurance. *Submitted.*

Lu, L. (2006). Some actuarial and statistical investigations into topics on genetics and insurance. PhD Thesis. *Heriot-Watt University, Edinburgh.*

Lu, L., Macdonald, A.S., Waters, H.R. and Yu, F. (2007). A dynamic family history model of hereditary nonpolyposis colorectal cancer and critical illness insurance. *To appear in Annals of Actuarial Science.*

Lynas, M. (1957). Dystrophia myotonica with special reference to Northern Ireland. *Annals of Human Genetics*, **21**, 318–351.

Lynch, H.T., Harris R.E., Guirgis, H.A., Maloney, K., Carmody, L.L. and Lynch, J.F. (1978). Familial association of breast/ovarian carcinoma. *Cancer*, **41**, 1543–1549.

Lynch, H.T. and Smyrk, T. (1996a). Hereditary Nonpolysis colorectal cancer (Lynch syndrome). *American Cancer Society*, **78**, 1149–1167.

Lynch, H.T. and Smyrk, T. (1996b). Colorectal cancer, survival advantage and hereditary nonpolyposis colorectal carcinoma. *Gastroenterology*, **110**, 943–954.

Lynch, H.T. and de la Chapelle, A. (2003). Hereditary colorectal cancer. *The New England Journal of Medicine*, **348**, 919–932.

Maas, O. (1937). Observations on Dystrophia Myotonica. *Brain*, **60**, 498–524.

MacCalman, L. (2009). Effects of genetic testing on insurance: Pedigree analysis and ascertainment adjustment. PhD Thesis. *Heriot-Watt University*.

Macdonald, A.S. (1997). How will improved forecasts of individual lifetimes affect underwriting?. *British Actuarial Journal*, **3**, 1044–1058.

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

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

Macdonald, A.S. (2003a). Genetics and insurance: What we have learned so far?. *Scandinavian Actuarial Journal*, **2003**, 324–328.

Macdonald, A.S. (2003b). Moratoria on the use of genetic tests and family history for mortgage-related life insurance. *British Actuarial Journal*, **9**, 279–311.

Macdonald, A.S., Waters, H.R. and Wekwete, C.T. (2003a). The genetics of

breast and ovarian Cancer I: A model of family history. *Scandinavian Actuarial Journal*, **1**, 1–27.

Macdonald, A.S., Waters, H.R. and Wekwete, C.T. (2003b). The genetics of breast and ovarian Cancer II: A model of critical insurance. *Scandinavian Actuarial Journal*, **1**, 28–50.

Macdonald, A.S. & McIvor, K.R. (2009). Modelling adverse selection in the presence of a common genetic disorder: The breast cancer polygene. *To appear in ASTIN Bulletin.*

Macdonald, A.S. and Tapadar, P. (2010). Multifactorial genetic disorders and adverse selection: Epidemiology meets economics. *To appear in Journal of Risk and Insurance.*

Mackay, G. (1993). Permanent health insurance: Overviews and market conditions in the U.K.. *Insurance: Mathematics and Economics*, **13**, 123–130.

MacMillan, J.C., and Harper, P.S. (1991). Single-gene neurological disorders in South Wales: an epidemiological study. *Ann Neurol*, **30**, 411–414.

Magee, A. (1996). The epidemiology and genetics of myotonic dystrophy in Northern Ireland. PhD Thesis. *Queen's University of Belfast.*

Magee, A. and Nevin, N.C. (1999). The epidemiology of myotonic dystrophy in Northern Ireland. *Community Genetics*, **2**, 179–183.

Mandel, J.S., Bond, J.H., Church, T.R., Snover, D.C., Bradley, G.M., Schuman, L.M. and Ederer, F. (1993). Reducing mortality from colorectal cancer by screening for recal occult blood. *The New England Journal of Medicine*, **328**, 1365-1371.

Mathieu, J., Braekeleer, M. and Prévost, C. (1990). Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). *Neurology*, **40**, 839–842.

Mathieu, J., De Braekeleer, M., Prévost, C. and Boily, C. (1992). Myotonic dystrophy: clinical assessment of muscular disability in an isolated population with presumed homogeneous mutation. *Neurology*, **42**, 203–208.

Mathieu, J., Allard, P., Potvin, L., Prévost, C. and Bégin, P. (1999). A 10-year study of mortality in a cohort of patients with myotonic dystrophy. *Neurology*, **52**, 1658–1662.

Medica, I., Marković, D., and Peterlin, B. (1997). Genetic epidemiology of myotonic dystrophy in Istria, Croatia. *Acta Neurol Scand*, **95(3)**, 164–166.

Mladenovic, J. (2003). Epidemiology of myotonic dystrophy, type 1 in central Serbia. MSc Thesis. *University of Belgrade, Belgrade.*University of Belgrade

Mladenovic, J., Pekmezovic, T., Todorovic, S., Rakocevic-Stojanovic, V., Romac, S. and Apostolski, S. (2005). Epidemiology of myotonic dystrophy type 1 in the population of central Serbia. *Vojnosanitetski pregled*, **62**, 377–382.

Mladenovic, J., Pekmezovic, T., Todorovic, S., Rakocevic-Stojanovic, V., Savic, D., Romac, S., and Apostolski, S. (2006). Survival and mortality of myotonic dystrophy type 1 (Steinert's disease) in the population of Belgrade. *European Journal of Neurology*, **13**, 451–454.

Mor-Cohen, R., Magal, N., Gadoth, N., Achiron, A., Shohat, T. and Shohat, M. (1997). The lower incidence of myotonic dystrophy in Ashkenazic Jews compared to North African Jews is associated with a significantly lower number of CTG trinucleotide repeats. *Isr J Med Sci*, **33**, 190–193.

Mori, M., Harabuchi, I., Hirotsugu, M., Casagrande, J.T., Henderson, B.E. and Ross, R.K. (1988). Reproductive, genetic, and dietary risk factors for ovarian cancer. *Am J epidemiol*, **128**, 771–777.

Morice, P., Camatte, S., Wicart-Poque, F., Atallah, D., Rouzier, R., Pautier, P., Pomel, C., Lhommé, C., Duvillard, P. and Castaigne, D. (2001).

Results of conservative management of epithelial malignant and borderline ovarian tumours. *Human reproduction update*, **9**, 185–192.

Morrison, P.J., Hodgson, S.V. and Haites, N.E. (2002). *Familial breast and ovarian cancer: Genetics, screening and management*. Cambridge University Press.

Mostacciulol, M.L., Barbujani, G., Armani, M., Danieli, G.A. and Angelini, C. (1987). Genetic epidemiology of myotonic dystrophy. *Genetic Epidemiology*, **4**, 289–298.

Nesterov, L.N., Sushcheva, G.P., Viatkina, S., and Nesterova, I. (1983). Myotonic Dystrophy. *Zh Nevropatol Psikhiatr Im S S Korsakova*, **83**, 1634–1641.

Novelli, G., Gennarelli, M., Menegazzo, E., Mostacciuolo, M.L., Pizzuti, A., Fattorini, C., Tessarolo, D., Tomelleri, G., Giacanelli, M., Danieli, G.A., Rizzuto, N., Caskey, C.T., Angelini, C. and Dallapiccola, B. (1993). $(CTG)_n$ triple mutation and phenotype manifestations in myotonic dystrophy patients. *Biochemical Medicine and Metabolic Biology*, **50**, 85–92.

O'Hoy, K.L., Tsilfidis, C., Mahadevan, M.S., Neville, C.E., Barceló, J., Hunter, A. and Korneluk, R.G. (1993). Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science*, **259**, 809–811.

O.N.S. (1997). *1990–1992 English Life Tables No. 15, Series DS No. 14*. Office for National Statistics.

O.N.S. (1999). *Cancer 1971–1997 CD-ROM*. Office of National Statistics, London.

O.P.C.S. (1991). *1990 mortality statistics: Cause. Series DH 12 No.17*. Office of Population Censuses and Surveys.

O.P.C.S. (1993). *1991 mortality statistics: Cause. Series DH 12 No.18*. Office of Population Censuses and Surveys.

Osame, M. and Furusho, T. (1983). Genetic epidemiology of myotonic dystrophy in Kagoshima and Okinawa districts in Japan. *Rinsho Shinkeigaku*, **23**, 1067–1071.

Papadopoulos, N. (2004). Microsatellite instability (MSI) in non-colonic, non-HNPCC tumors: 'Unstable' evidence?. *Annals of Oncology*, **10**, 751–752.

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

Percesepe, A., Benatti, P., Roncucci, L., Sassatelli, R., Fante, R., Ganazzi, D., Bellacosa, A., Genuardi, M., Neri, G., Viel, A. and Ponz, D.L.M. (1997). Survival analysis in families affected by hereditary non-polyposis colorectal cancer. *International Journal of Cancer*, **71**, 373–376.

Pfisterer, J., Plante, M., Vergote, I., du Bois, A., Hirte, H., Lacave, A.G., Wagner, U., Staähle, Stuart, G., Kimmig, R., Olbricht, S., Le, T., Emerich, J., Kuhn, W., Bentley, J., Jackisch, C., Lück, H., Rochon, J., Zimmermann, A.H. and Eisenhauer, E. (2006). Gemcitabine plus carboplatin compared with carboplatin in patients with platinum-sensitive recurrent ovarian cancer: An intergroup trial of the AGE-OVAR, the NCIC CTG, and the EORTC GCG. *Journal of Clinical Oncology*, **24**, 4699–4707.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1988). *Numerical recipes in C: The art of scientific computing*. Cambridge University Press.

Ravine, D., Walker, R., Gibson, R., Forrest, S., Richards, R., Friend, K., Sheffield, L., Kincaid-Smith, P. & Danks, D. (1992). Phenotype and genotype heterogeneity in autosomal dominant polycystic kidney disease.. *The Lancet*, **340**, 1330–1333.

Renkonen-Sinisalo, L., Aarnio, M., Mecklin, J.P. and Järvinen, H.J. (2000). Surveillance improves survival of colorectal cancer in patients with hereditary non-polyposis colorectal cancer. *Cancer Detect Prev*, **24**, 137–142.

Reznek, R.H. (2007). *Cancer of the ovary*. Cambridge University Press.

Risch, H.A., Marrett, L.D. and Howe, G.R. (1994). Parity, contraception, infer-

tility, and the risk of epithelial ovarian cancer. *Am J Epidemiol*, **140**, 585–597.

Rodriguez-Bigas, M., Boland, C. and Hamilton, S. (1997). A National Cancer Institute workshop on hereditary non-polyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. *Journal of the National Cancer Institute*, **89**, 1758-1762.

Sanders, A.J. & Silby, N.F. (1988). Actuarial aspects of PHI in the U.K.. *Journal of the Institute of Actuaries Students' Society*, **31**, 1–57.

Sankila, R., Aaltonen, L.A., Jarvinen, H.J. and Mecklin, J.P. (1996). Better survival rates in patients with MLH1-associated hereditary colorectal cancer. *Gastroenterology*, **110**, 682–687.

Schilder, J.M., Thompson, A.M., DePriest, P.D., Ueland, F.R., Cibull, M.L., Kryscio, R.J., Modesitt, S.C., Lu, K.H., Geisler, J.P., Higgins, R.V., Magtibay, P.M., Cohn, D.E., Powell, M.A., Chu, C., Stehman, F.B. and van Nagell, J. (2002). Outcome of reproductive age women with stage IA or IC invasive epithelial ovarian cancer treated with fertility-sparing therapy. *Gynecologic oncology*, **87**, 1–7.

Schulmann, K., Brasch, F.E., Kunstmann, E., Engel, C., Pagenstecher, C., Vogelsang, H., Krüger, S., Vogel, T., Knaebel, H.P., Rüschoff, J., Hahn, S.A., Knebel-Doeberitz, M.V., Moeslein, G., Meltzer, S.J., Schackert, H.K., Tympner, C., Mangold, E., Schmiegel, W. and The German HNPCC Consortium (2005). HNPCC-associated small bowel cancer: clinical and molecular characteristics. *Gastroenterology*, **128**, 590–599.

Shamsunder, S., Kumar, L., Gupta, S., Kumar, S., Bhatla, N., Singh, R. and Kochupillai, V. (2000). Chemotherapy in recurrent epithelial ovarian cancer (EOC): An analysis of prognostic factors. *Journal of Obstetrics and Gynaecology Research*, **3**, 215–222.

Shelbourne, P., Davies, J., Buxton, J., Anvret, M., Blennow, E., Bonduelle, M., Schmedding, E., Glass, I., Lindenbaum, R., Lane, R., *et al.* (1993). Direct diagnosis of myotonic dystrophy with a disease-specific DNA marker. *N Engl J Med*, **328(7)**, 471–475.

Shepherd, J.H. (1989). Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol*, **96(8)**, 889–892.

Siciliano, G., Manca, M., Gennarelli, M., Angelini, C., Rocchi, A., Iudice, A., Miorin, M. and Mostacciulol, M. (2001). Epidemiology of myotonic dystrophy in Italy: re-apprisal after genetic diagnosis. *Clinical Genetics*, **59**, 344–349.

Sizhong, Z., Hui, W., Agen, P., Cuiying, X., Ge, Z., Yiping, H., and Jiayou, C. (2000). Low incidence of myotonic dystrophy in Chinese Hans is associated with a lower number of CTG trinucleotide repeats. *Am J Med Genet*, **96**, 425–428.

Thomas, G.M. (1994). Radiotherapy in early ovarian cancer. *Gynecologic Oncology*, **55**, S73-S79.

Thomasen, E. (1948). *Myotonia*. Universitetsforlaget.

Tomoda, H., Baba, H. and Akazawa, K. (1999). Prolonged survival in hereditary nonpolyposis colorectal cancer. *Oncology reports*, **6**, 321–324.

Trimbos, J.B., Vergote, I., Bolis, G., Vermorken, J.B., Mangioni, C., Madronal, C., Franchi, M., Tateo, S., Zanetta, G., Scarfone, G., Giurgea, L., Timmers, P., Coens, C., Pecorelli, S., EORTC-ACTION collaborators. European Organization for Research and Treatment of Cancer-Adjuvant Chemotherapy in Ovarian Neoplasm (2003). Impact of adjuvant chemotherapy and surgical staging in early-stage ovarian carcinoma: European Organisation for Research and Treatment of Cancer-Adjuvant ChemoTherapy in Ovarian Neoplasm trial. *Journal of the National Institute*, **95**, 113–125.

Tsakiris, D., Jones, E., Brigs, D., Elinder, C., Mehls, O., Mendel, S., Pic-

coli, G., Rigden, S., dos Santos, J., Simpson, K. and Vanrenterghem, Y. (1999). Deaths Adult Polycystic Kidney Disease and Insurance: Genetic Heterogeneity 37 within 90 days from starting renal replacement therapy in the ERA-EDTA registry between 1990 and 1992. *Nephrology Dialysis Transplantation*, **14**, 2343–2350.

Vasen, H.F., Mecklin, J.P., Watson, P., Utsunomiya, J., Bertario, L., Lynch, P., Svendsen, L.B., Cristofaro, G., Müller, H., Khan, P.M. *et al.* (1993). Surveillance in hereditary nonpolyposis colorectal cancer: an international cooperative study of 165 families. The International Collaborative Group on HNPCC. *Dis Colon Rectum*, **36**, 1–4.

Vasen, H., Stormorken, A., Menko, F., Nagengast, F., Kleibeuker, J., Griffioen, G. and Taal, B. (2001). MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: A study of hereditary nonpolyposis colorectal cancer families. *Jounal of Clinical Oncology*, **19**, 4074–4080.

Vergote, I., Kærn, J., Abeler, V.M., Pettersen, E.O., De Vos, L.N. and Tropé, C.G. (1993). Analysis of prognostic factors in stage I epithelial ovarian carcinoma: Important of degree of differentiation and deoxyribonucleic acid ploidy in predicting relapse. *American Journal of Obstetrics & Gynecology*, **169**, 40–52.

Vergote, I., De Brabanter, J., Fyles, A., Bertelsen, K., Einhorn, N., Sevelda, P., Gore, M.E., Kaern, J., Verrelst, H., Sjövall, K., Timmerman, D., Vandewalle, J., Van Gramberen, M. and Tropé, C.G. (2001). Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet*, **357**, 176–182.

Vergote, I. and Trimbos, B.J. (2003). Treatment of patients with early epithelial ovarian cancer. *Current Opinion in Oncology*, **15**, 452-455.

Watson, P., Lin, K.M., Rodriguez-Bigas, M.A., Smyrk, T., Lemon, S., Shashidharan, M., Franklin, B., Karr, B., Thorson, A. and Lynch, H.T. (1998).

Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. *Cancer*, **83**, 259–266.

Wekwete, C.T. (2002). Genetics and critial illness insurance underwriting: Models for breast cancer and ovarian cancer and for coronary heart disease and stroke. PhD Thesis. *Heriot-Watt University.*

Wehner, A.P. (1994). Biological effects of cosmetic talc. *Food Chem Toxicol*, **32**, 1173–1184.

Wendt, G.G. and Drohm, C. (1972). Die Huntingtonsche chorea. Eine populations gentische studie. *Advances in Human Genetics*, **IV**, 1–121.

The WHO Collaborative Study of Neoplasia and Steroid Contraceptives (1989). Epithelial ovarian cancer and combined oral contraceptives. *Int J Epidemiol*, **18**, 538–545.

Wilkie, A.D. (2000). Report by the independent actuary on the application by the Association of British Insurers to the Genetics and Insurance Committee for approval to use genetic test results for insurance risk assessment — use of Huntington's disease test in life insurance. *Genetics and Insurance Committee, Department of Health.*

Zanetta, G., Chiari, S., Rota, S., Bratina, G., Maneo, A., Torri, V., Mangioni, C. (1997). Conservative surgery for stage I ovarian carcinoma in women of childbearing age. *British journal of obstetrics and gynaecology*, **104**, 1030–1035.

Zanetta, G., Rota, S., Chiari, S., Bonazzi, C., Bratina, G., Torri, V. and Mangioni, C. (1998). The accuracy of staging: An important prognostic determinator in stage I ovarian carcinoma. *Annals of Oncology*, **9**, 1097–1101.

Zhang, H., Kavanagh, J.J., Hu, W., Liao, Q. and Fu, S. (2007). Hormonal therapy in ovarian cancer. *International journal of gynecological cancer*, **17**, 325-338.