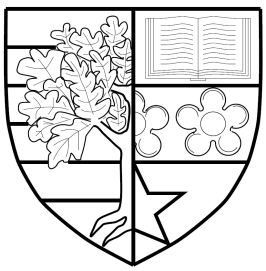
GENETICS, INSURANCE, AND CARDIOMYOPATHIES: A CASE STUDY OF HYPERTROPHIC CARDIOMYOPATHY

by

Oytun Haçarız



Submitted for the degree of
Doctor of Philosophy

Department of Actuarial Mathematics and Statistics
School of Mathematical and Computer Sciences
Heriot-Watt University

March 2020

Abstract

The economic impact of genetic information on life insurance has been discussed since DNA-based genetic testing became available in the 1990s. Macdonald & Yu (2011) estimated the highest increases in life insurance premium rates were about 0.6% if genetic test results were undisclosed to the insurers. Howard (2014) concluded that premium increases could be as high as 12% if the insurers were unable to access genetic test results. Although these two studies used different methodologies, the differences in their conclusions were due to the inclusion of cardiomyopathies (inherited heart muscle disorders), which were absent in the first of these studies. Hypertrophic Cardiomyopathy (HCM) is the most common of these disorders with a prevalence rate estimated to be 0.2% in the general population.

We identify a mathematical model of the impact of genetic testing in HCM in a life insurance market under adverse selection. Then, we estimate the necessary premium increases to meet adverse selection costs and survey significant factors leading to increases and decreases in adverse selection costs. A novel feature of our model is that it includes 'cascade genetic testing', which is the form of genetic testing that is the most associated with HCM, in nuclear families.

We conclude that the range of possible adverse selection costs is large, but the costs with the most reasonable assumptions are small and consistent with Macdonald & Yu (2011). Much higher costs depend on 'adverse selectors' treating life insurance as a financial investment and taking out extremely large sums insured, and also disregard selection and ascertainment biases in the epidemiological literature.

Acknowledgements

I am greatly thankful to both of my supervisors, Professor Angus S. Macdonald and Doctor Torsten Kleinow, for their extraordinary supervision during my PhD studies. Undoubtedly, without their directions and encouragement, it is uncertain how much of this thesis would exist. Especially, very special thanks to Professor Angus S. Macdonald for being so patient, by giving very good feedback on my writing and replying my endless questions, during this time. Additionally, I appreciate the helpful comments of the late Professor A. J. (Sandy) Raeburn and Doctor Wayne Lam in terms of the clinical perspective of Hypertrophic Cardiomyopathy (HCM).

I would like to thank all my family, especially my parents, brother, and maternal aunts, very much for their constant financial and emotional support through this process, and always being the main characters in any achievement in my life.

I am also very thankful to all my friends for morale support in the meantime; in particular, Ana, Andres, Ayşe, Bemsibom, Bernhardt, Cristian, Francesco, Carsten, George, Chaofan, Giovanni, Hove, Jeta, and Matthias.

One final but the deepest thanks to the Ministry of National Education of the Republic of Turkey for funding my PhD (including MSc) and living expenses abroad.

Contents

List of Tables

9.9 Percentage increases in premium rates in respect to a sensitivity analysis of fatal HCM associated with that the annual hazard rate of non-fatal HCM is zero at all ages in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu_{x,z}^{i02}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ_x^{ijg}: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ_x^{ij8}: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ_x^{ij8}: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.182

List of Figures

Chapter 1

Introduction

1.1 Genetics, Insurance and Cardiomyopathies

Genetics is a field of science specializing in the study of genes. A gene in humans, a particular region of DNA (deoxyribonucleic acid) inherited from parents, is involved in the generation of the physical traits (phenotype) of the offspring. It is estimated that humans have about 30,000–35,000 genes. A gene mutation is a permanent alteration within any gene sequence, which might cause a 'genetic disorder'. There are several types of genetic disorders which are classified as follows (see Sudbery (2002)):

- **Single-Gene Disorders:** These are genetic disorders caused by an alteration in a single gene. Their inheritance pattern is governed by Mendel's law of genetics. Therefore, they are also called 'Mendelian' disorders. They can be inherited as recessive, both copies of the mutant gene are necessary, or dominant, one copy of the mutant gene is sufficient. As a result, the risk of a family member carrying a mutation can be estimated by tracking family history. Their manifestation (onset) might be at earlier stages of life, such as Cystic Fibrosis, or at later stages of life, such as Huntington's Disease. They are relatively rare in the general population, but their impact on life expectancy is significant.

- **Multifactorial (Complex) Disorders**: These are genetic disorders caused

by a composition of many mutant genes along with environmental and lifestyle factors. They do not follow a clear pattern of Mendelian inheritance. The environmental and lifestyle changes might even decrease the risk of disorders. Various cancers, diabetes, migraine, and asthma can be given as common examples of these disorders.

- **Chromosomal Disorders**: These are genetic disorders caused by an alteration of the genetic structure of chromosomes or the presence of an extra chromosome. The most prevalent of these disorders is Down's syndrome. The majority of chromosomal disorders are not inherited through families.

- **Mitochondrial Disorders**: These are genetic disorders caused by an alteration of the genetic structure of the mitochondria which only mothers can pass to their offspring. They also do not follow a typical pattern of Mendelian inheritance.

Until DNA-based genetic testing began to become available in the 1990s, many of these genetic disorders were studied through family histories. Genetic testing enabled individuals to learn for certain whether or not they are affected by the mutations present in their families. Genetic testing also helped individuals to diagnose gene mutations and start their treatment. Even if there is no certain cure for the most genetic disorders, early diagnosis is important for early clinical care.

On the other hand, individuals and patient groups have often been concerned that genetic test results would be so highly predictive of greatly increased risk that insurance would be denied or become unaffordable to some. They have often advocated banning the use of genetic test results by insurers. To protect individuals against 'genetic discrimination', insurers are now banned from using genetic test results in underwriting in many countries. Different countries have different practices to protect individuals against 'genetic discrimination' as follows:

(a) In the UK, the Association of British Insurers (ABI) has currently a voluntary agreement with the government on the use of genetic test results in underwriting insurance policies. According to this agreement:

(i) Genetic test results are classified as 'predictive', meaning that they allow the prediction of future risk of developing genetic disorders in healthy gene mutation carriers, or 'diagnostic', confirming the clinical manifestation of genetic disorders.

(ii) Insurers are allowed to use 'diagnostic' genetic test results, but they are not allowed to use 'predictive' genetic test results except for a sum assured (per individual) of above

- £500,000 in life insurance (an exception of Huntington's disease),
- £300,000 in critical illness insurance, and
- £30,000 per annum in income protection insurance.

The only test that may be used to date for over £500,000 of life insurance has been in the case of Huntington's disease.

(iii) Individuals, however, can disclose favourable predictive genetic test results, such as negative test results.

(iv) Moreover, if genetic test results, accidentally or voluntarily, are shared by individuals, insurers can only use favourable genetic test results in underwriting. In this situation, the agreement obliges insurers to avoid the use of unfavourable genetic test results in underwriting as long as individuals do not ask for a sum assured more than the levels in point (ii) above.

(b) In Canada, the law called the Genetic Non-Discrimination Act (GNDA), previously referred as Bill S-201, bans insurers from using all (including predictive and diagnostic) genetic test results. However, the law does not specifically refer to insurers, and does not define an exception for high levels of sum assured as in the UK's agreement.

(c) In the USA, the law called the Genetic Information Nondiscrimination Act (GINA) bans only health insurers from using genetic information, including genetic tests and family history, in underwriting. However, in the law, no restrictions are directed to life insurance.

See Prince (2019) for the regulations of the countries above (and more) in detail.

As a result of such regulations, insurers have been concerned that if individuals knew of a genetic test result indicating greatly elevated risk, and were not obliged to disclose this to an insurer, they would be able to obtain insurance cover for much below its true cost — classical adverse selection.

Attention was focused on a small subset of genetic disorders deemed to expose insurers to particularly high risks of adverse selection. These were the single-gene disorders caused by the mutations in a single gene, and which were dominantly inherited (meaning that one copy of the mutation inherited from either parent could lead to the disorder). Among a subset of these disorders, called 'late-onset disorders', onset of symptoms is usually delayed into adulthood. Carriers of such mutations could:

- remain completely free of symptoms until early adulthood;
- buy life and health insurance during this time; and
- develop symptoms and/or die while an insurance policy was still in force.

Even in the absence of genetic tests, dominantly inherited disorders revealed their presence by the pattern of inheritance in families. Thus they had been studied by epidemiologists for many years. Insurers, by asking questions during underwriting about an applicant's family history, could also learn about the applicant's risk of carrying a mutation. (Insurers' questions were normally confined to first-degree relatives; that is, parents and siblings of the applicant.) So, the advent of genetic tests did not expose insurers to completely new risks. A family history might reveal that an applicant had a 50% chance of having inherited the mutation. A genetic test would reveal for certain whether they had or had not.

Late-onset disorders are, fortunately, quite rare. For this reason, attempts to model the potential costs of adverse selection, in terms of across-the-board increases to premiums, found that they would be quite small. Macdonald & Yu (2011), modelled six major single-gene disorders, based on published epidemiology, and estimated the cost of adverse selection under the following three forms of moratorium

on insurers' use of genetic test results.

- Insurers may not use any genetic test results.
- Insurers may not use adverse genetic test results, but may use test results that show a mutation to be absent. (This reflects the practice in the UK, under the moratorium agreed between the government and the insurance industry.)
- Insurers are banned from using family history as well as genetic test results.

Using a variety of scenarios, the highest increases in life insurance premium rates were about 0.6% if insurers could use family history, and about 1% if they could not. These assumed that persons buying insurance after an adverse test result did not buy above average amounts of insurance, but if they did the premium increases would change *pro rata*.

Absent from the major classes of single-gene disorders modelled by Macdonald & Yu (2011) were inherited disorders of the heart muscle, known collectively as *cardiomyopathies*. Howard (2014), in a report commissioned by the Canadian Institute of Actuaries, proposed a model which included the main cardiomyopathies. He concluded that, if genetic test results were not disclosed to Canadian insurers:

- premium increases caused by adverse selection could be as high as 12%; and
- the overall mortality experience of Canadian insurers could increase by 36% for males and 58% for females.

Although Howard (2014) and Macdonald & Yu (2011) used different methodologies, it was clear that the differences in their conclusions were in part due to the inclusion of the cardiomyopathies in one model only.

The purpose of this study is to develop a mathematical model of the most prevalent cardiomyopathy (Hypertrophic Cardiomyopathy (HCM)) and model the possible adverse selection costs in life insurance.

1.2 Healthy Heart and Cardiomoypathies

1.2.1 Major Features of a Healthy Heart

We aim to show the major features of a healthy heart to improve the understanding of disorders of the heart. Therefore, this section describes the basics of a healthy heart, which can be accessed from many sources, for example the booklets of the Heart Societies such as American Heart Association (AHA), British Heart Foundation (BHF), etc. We refer in particular to Seidman & Seidman (2001), Whitaker (2006), AHA (2011), Khurana (2014), and McKenna & Elliott (2015).

The heart is a muscular pump which is responsible for the circulation of the blood to the lungs and the body. The heart is separated into four chambers; two of them are located on the right of the heart as the right atrium (upper-right) and right ventricle (lower-right), and the other two are located on the left of the heart as the left atrium (upper-left) and left ventricle (lower-left). These right and left chambers are divided by the septum wall.

The heart includes four valves. Two of them help the blood flow into the ventricles from the atriums, namely the tricuspid valve (between the right atrium and right ventricle), and the mitral valve (between the left atrium and left ventricle). The other two control the blood flow out of the ventricles, namely the pulmonary valve (between the right ventricle and pulmonary artery), and the aortic valve (between the left ventricle and aorta).

The healthy heart pumps the blood in the pulmonary and systemic circulations. Deoxygenated blood from the body reaches the right atrium via the superior and inferior vena cava veins. Throughout diastole (cardiac relaxation), the blood passes into the right ventricle via the tricuspid valve. In pulmonary circulation, throughout systole (cardiac contraction), the blood in the right ventricle is moved into the lungs to be oxygenated, through the pulmonary valve and pulmonary artery. The oxygenated blood from the lungs reaches the left atrium via the pulmonary vein. And, throughout diastole (cardiac relaxation), the oxygenated blood in the left atrium enters the left ventricle via mitral valve. In systemic circulation, throughout

systole (cardiac contraction) the oxygenated blood in the left ventricle is pumped to the body, through the aortic valve into the aorta.

The contraction of the heart muscle is regulated by electric activity via signals between specialized heart muscle fibres, which is called 'the conduction system of the heart'.

The wall structure of the heart consists of three layers; the endocardium, the myocardium (cardiac or heart muscle), and the pericardium. The endocardium is the thin layer located inside the heart which covers the chambers and valves. The pericardium is the outer layer that holds and protects the heart. The myocardium, between the endocardium and pericardium, is the muscular layer of the heart responsible for pumping the blood out of the heart.

1.2.2 Major Features of Cardiomyopathies

Cardiomyopathies are disorders of the heart affecting the myocardium. They are responsible for sudden, unexpected, heart attacks in otherwise healthy young people, widely reported when they happen to professional sports stars. They are similar to the 'classical' single-gene disorders modelled in Macdonald & Yu (2011), in that they are single-gene, dominantly inherited Mendelian disorders, but there are also important differences.

(a) Taking inherited breast cancer as an example, a genetic test for BRCA1 or BRCA2 gene mutations may reveal the increased risk, while there are no cancerous or precancerous tissues in the body. And, if these ever do appear, it may be several decades after the genetic test was taken. In insurance terms, a test for mutations in these genes is predictive, not diagnostic. Note that the moratorium in the UK bans insurers from using predictive tests but not from using diagnostic tests.

(b) The changes to the heart muscle associated with cardiomyopathies are often, but not always, present by adolescence or even earlier. They are, in principle, capable of being detected by non-genetic clinical tests, such as electrocardiogram (ECG) and echocardiography. In insurance terms, a genetic test for a

cardiomyopathy would seem often to be diagnostic, because it reveals a pre-existing condition; the heart muscle is already affected. However, some may argue that genetic tests for cardiomyopathies are diagnostic of the physiological change, but predictive of the increased risk of death.

1.2.3 Classification of Cardiomyopathies

The term 'cardiomyopathy' covers a wide range of disorders, each of which may be associated with mutations in more than one gene. Some disorders present as changes to the musculature of the heart. Others present as disruption to the electrical signals controlling the heart's rhythm (ion channelopathies). Thus they present a significantly greater modelling challenge than many of the classical single-gene disorders.

We follow two major reports in the classifications below. The European Society of Cardiology (ESC) Report (Elliott et al. 2008) classified the disorders as;

- Hypertrophic Cardiomyopathy (HCM),
- Dilated Cardiomyopathy (DCM),
- Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC),
- Restrictive Cardiomyopathy,
- Unclassified.

The American Heart Association (AHA) Scientific Statement (Maron et al. 2006) further classified ion channelopathies as:

- Long-QT Syndrome (LQTS),
- Brugada Syndrome,
- Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT),
- Short-QT Syndrome (SQTS),
- Asian Sudden Unexplained Nocturnal Death Syndrome (SUNDS).

Hypertrophic Cardiomyopathy (HCM) is regarded to be the most prevalent of these disorders in the general population, (Maron et al. 2014), and is the subject of this thesis.

1.3 Major Features of Howard (2014)

Howard (2014) models genetic disorders under adverse selection in the life insurance market in Canada with regard to the law (GNDA, see Section 1.1), prohibiting insurers to access genetic test results. The model includes thirteen genetic disorders (including HCM) which are regarded as being significant in giving rise to adverse selection costs once underwriters are not able to underwrite the full risk of the individuals affected by these disorders. Howard (2014) presented two conclusions:

(a) the benefit claim costs in Canada under adverse selection would be of the order of 10% of the total benefit claim costs in one year, and

(b) the overall mortality experience under adverse selection would increase by about 40%.

HCM was the most prevalent and the second most expensive disorder in Howard (2014). We wish to understand why HCM is so expensive compared to the other disorders in the same model. The fundamental assumptions about HCM in Howard (2014) (see Section 10.2.1) are: the prevalence of HCM mutations in the general population is 0.2%; and the annual mortality rate of HCM is 0.01, or $q_x = 0.01$ (Section 3.9.8). These rates seem to be widely cited in the epidemiological literature, but they should be evaluated with care, see Section 1.5, and Chapters 2 and 3.

Another key assumption is that adverse selectors take out sums assured of $1,000,000 (ten times the assumed normal sum assured of $100,000). There is almost no evidence of what would be a realistic assumption of what sums assured adverse selectors would purchase; in Sections 9.7, 11.3, and 11.4, we discuss this.

We will go into detail on Howard (2014) in Chapter 10 after we have defined and described our model with corresponding adverse selection costs.

1.4 Multiple-State Models

Multiple-state models have been used to model genetic disorders in health and life insurance under adverse selection have been used in many studies such as Macdonald (1999), Subramanian et al. (1999), Macdonald et al. (2003*a*), Macdonald et al. (2003*b*), Gutiérrez & Macdonald (2004), Gui et al. (2006), Lu et al. (2007), and Macdonald & Yu (2011). They provide a very flexible approach to representing transitions between states, and are often used to model transitions between states of health (see Dickson et al. (2013) and Macdonald et al. (2018)). Other examples of transitions can be a transition from an untested to a tested state, representing the event of genetic testing, or an uninsured to an insured state, representing the event of purchasing insurance.

Following this pattern, in this study, we will use multiple-state models to model genetic testing in HCM and its impact on life insurance under adverse selection. To the best of our knowledge, this study contributes a novel feature to the published studies above, namely modelling cascade genetic testing within families (Chapters 4 and 5).

1.5 Motivation

The motivation for a mathematical model of HCM for life insurance might come from many directions, but ours mainly comes from adverse selection. Adverse selection in insurance is the consequence of asymmetry of information between insurers and individuals. Insurers may lack information known to the applicant, possibly because of prohibition on asking the applicant for the information.

Adverse selection in genetics and insurance is a good example of adverse selection in insurance because in many countries (including the UK) genetic test results are not disclosed to insurers (Section 1.1). Epidemiological studies allow us to understand the risks of genetic disorders when genetic test results are not disclosed to insurers. If insurers understood these epidemiological risks well, they could allow for them in pricing, and may not then suffer from adverse selection losses due to

unknown genetic test results. However, the epidemiology of genetic disorders is still evolving. For example, in HCM, two figures are widely cited: the population prevalence of HCM of 0.2% and the annual mortality rate (q_x) of HCM of 1%. Chapters 2 and 3 show that the first of these might be greatly underestimated, and the second greatly overestimated, with an impact on adverse selection costs (Chapter 9).

Even if the epidemiological data was thought to be mature, we still do not know what would be a realistic assumption for how much insurance or what amounts of sum assured, 'adverse selectors' would buy. Some might set these behavioural parameters in a conservative way (very large sum assured) which tends to make the (expected) adverse selection costs very large. But, how reasonable is this assumption? In Sections 9.7, 11.3, and 11.4, we discuss the results of published studies that might give some intuition about what would motivate the purchase behaviour of adverse selectors.

1.6 Plan of the Thesis

We identify a mathematical model of genetic testing in HCM and discover the possible adverse selection costs in a life insurance market.

In Chapter 2, we present the epidemiological features of HCM, such as its onset, the occurrence of major clinical events, and prevalence.

In Chapter 3, we develop a mathematical model of HCM which is a multiple-state model in a Markovian setting, and which we call the epidemiological model. The epidemiological model is parametrised by transition intensities. The occupancy probabilities in each model state can be obtained from these transition intensities by solving the Kolmogorov forward equations. We do not yet introduce states and transitions representing genetic testing because of the nature of genetic testing in HCM, namely cascade genetic testing. Cascade genetic testing cannot easily be modelled in a Markovian setting because the testing behaviour depends on not only the currently occupied state, but also family history.

Thus, in Chapter 4, we describe cascade genetic testing in HCM. We parametrise the uptake rate of cascade genetic testing within the families affected by HCM.

In Chapter 5, we describe a simulation model of cascade genetic testing in HCM, which we call the testing model. It is built on the epidemiological model (adding states representing genetic testing). We simulate the testing behaviour of the individuals in HCM families. Unlike the epidemiological model, the testing model is not a Markov model because under cascade genetic testing, transition intensities from an untested to a tested state depend on the currently occupied state and family history. In theory, the state space of the model can be extended without limit to keep it Markov, but this becomes very unwieldy. Therefore, occupancy probabilities in tested states can no longer be obtained by solving the Kolmogorov equations. Instead, we compute these by simulating life histories in HCM families explicitly. The testing model is partly Markov because we assume that the individuals in non-HCM families never take up genetic testing and their life histories can still be modelled in a Markovian setting and the occupancy probabilities can be obtained for such individuals as described in Chapter 3. This leads us to divide the general population into two: HCM and non-HCM families.

In Chapter 6, we discuss life insurance mathematics, allowing us to extend our model by adding insurance purchase states and so to calculate insurance losses arising from persons in these insured states. Since this life insurance mathematics necessitates an elaborate notation, we present the details in Appendix A.

In Chapter 7, we add more states representing insurance purchase to the testing model, and obtain an adverse selection model of HCM for life insurance, which we call the adverse selection model.

In Chapter 8, we describe a combined numerical technique, using Thiele's differential equations and Monte-Carlo estimation, for calculating expected insurance losses from the individuals in non-HCM and HCM families, respectively. We define our measure of insurance losses under adverse selection, which will be the necessary premium increases to recoup adverse selection costs.

In Chapter 9, we provide our results based on the measure in Chapter 8. We survey extensively the factors that are significant for increasing and decreasing adverse selection costs.

In Chapter 10, we compare our study and results with those of Howard (2014), including the addition of lapse states to the adverse selection model, which we call the lapse model.

In Chapter 11, we discuss our results and make our conclusions.

Chapter 2

Hypertrophic Cardiomyopathy (HCM)

2.1 Introduction

In this chapter, we introduce the epidemiological features of Hypertrophic Cardiomyopathy (HCM) before modelling the epidemiology of HCM in Chapter 3.

In Section 2.2, we discuss the clinical features of HCM. In Section 2.3, we briefly describe the genetic substrate of HCM. In Section 2.4, we discuss HCM-related endpoints. In Section 2.5, we discuss the prevalence of HCM in the general population. In Section 2.6, we move from the epidemiology of HCM to a mathematical model of the epidemiology of HCM.

2.2 The Clinical Features of HCM

2.2.1 Onset

HCM is the thickening (hypertrophying) of the muscular wall of the left ventricle (LV) of the heart. In most cases, the thickening of the heart wall may have taken place by adolescence or early adulthood. In other cases, the process may be delayed until later ages. See Pokorski (2002). We call these early-onset and late-onset HCM respectively (though these terms are not used in the clinical literature). They are

Table 2.1: New York Heart Association (NYHA) Functional Classification.

NYHA Class	Limitations of Physical Activity	Symptoms of Heart Failure
I	None	None
II	Mild	Arise with ordinary physical activity (comfortable at rest)
III	Marked	Arise with less than ordinary physical activity (comfortable at rest)
IV	Severe	Arise even at rest

associated with mutations in different genes, see Sections 2.3.1.1 and 3.8.

2.2.2 Diagnosis

Diagnosis depends on having a clear definition of HCM. This is given by the American College of Cardiology Foundation (ACCF)/the American Heart Association (AHA) Guidelines (Gersh et al. 2011) and the European Society of Cardiology (ESC) Guidelines (Elliott et al. 2014), as follows.

(a) In adults Left Ventricular Wall Thickness (LVWT) generally greater than or equal to 15 mm. (The manifestation of LVWT at 13–15 mm can also be evaluated as the sign of HCM).

(b) In children, LVWT more than 2 standard deviations above the mean related to age, gender, or body structure.

For our purposes, these criteria divide the life history of a person with an HCM-related mutation into a period before onset and a period after onset, with age-at-onset defined as the earliest age at which the definition above is met. This allows for the possibility that onset never occurs, which it is convenient to represent as 'age-at-onset $= +\infty$'.

Clinical diagnosis is made by machine imaging techniques such as echocardiography and cardiac magnetic resonance (CMR). An electrocardiogram (ECG) is also recommended, especially for the first clinical diagnosis, to determine any rhythmical abnormalities.

2.2.3 Symptoms

While most persons with HCM do not develop any symptoms, a minority manifest the symptoms of chest pain, dyspnoea (shortness of breath), palpitations and syncope (fainting) (Elliott et al. 2014).

Any symptoms usually start many years after the clinical existence of an ECG abnormality or increased LVWT (Elliott et al. 2014). They might arise and be stable and of mild degree, or progress through to a severe degree, see Table 2.1. They can be relieved by drugs, but surgery, including heart transplantation, is recommended for HCM patients with drug resistant severe symptoms (Maron et al. 2014) (see Section 2.4.1).

2.3 The Genetics of HCM

HCM-related mutations occur in genes related to sarcomeres, proteins involved in contractions of the heart muscle. More than eight such genes are known (Gersh et al. 2011, Seidman & Seidman 2001). Mutations are dominantly inherited, with the following consequences.

(a) It is sufficient for one parent to pass the mutation to a child for that child to be affected. Since HCM-related mutations are moderately rare, we ignore the possibility that a person carries mutations in two genes.

(b) Each child of an affected parent will inherit the parent's mutation with probability 1/2, because of Mendel's laws.

2.3.1 Gene Mutations Associated with HCM

We follow Elliott et al. (2014) to define the type of HCM-related gene mutations and their prevalences in the HCM population, consisting of HCM-related mutation carriers in the general population.

2.3.1.1 Known Gene Mutations Associated with HCM

Between 40% and 60% of individuals with HCM are found to carry a mutation in a known sarcomere-related gene (Elliott et al. 2014). Mutations are found most frequently in the MYBPC3 gene (15–30%) and MYH7 gene (10–20%), and less frequently in the TNNT2 gene (3–5%), TNNI3 gene (<5%) and TPM1 gene (<5%). Mutations in each gene are heterogeneous; that is, not confined to a single location on the gene.

Mutations in the MYBPC3 gene are associated with late-onset HCM while mutations in other genes are associated with early-onset HCM (Niimura et al. 1998, Pokorski 2002, Jensen et al. 2013) (see Section 2.2.1).

In our study, therefore, we divide carriers of known HCM-related mutations into two sub-populations:

(a) a known early-onset HCM mutation carrier sub-population,

(b) a known late-onset HCM mutation carrier sub-population.

2.3.1.2 Unknown Gene Mutations Associated with HCM

Approximately between 25% and 30% of individuals with HCM are found not to carry a known HCM-related mutation (Elliott et al. 2014). This means that genetic testing, either does not detect a known HCM-related mutation, or finds a variant of unknown significance (VUS).

Our presumption is that a clinically affected person who does not test positive for a known HCM-related mutation must carry a mutation that has yet to be identified.

Then, we obtain another two sub-populations for the carriers of unknown mutations:

(a) an unknown early-onset HCM mutation carrier sub-population,

(b) an unknown late-onset HCM mutation carrier sub-population.

In total, therefore, the population of HCM mutation carriers is divided into four sub-populations, according as the mutation is known/early-onset, unknown/early-onset, known/late-onset and unknown/late-onset.

2.3.1.3 Other Disorders Associated with HCM

Approximately between 5% and 10% of individuals with HCM are associated with other disorders, of which, most are genetic disorders while some are non-genetic disorders (Elliott et al. 2014). Gersh et al. (2011) describe them as a 'clinical mimic' of HCM.

For our purposes, we are less interested in the 'other genetic disorders associated with HCM'. They are complex multi-system disorders in which the genetic component is usually autosomal recessive, and are most likely to affect infants, children and adolescents:

(a) They manifest themselves in the early years of life, so they are of no interest for life insurance questions.

(b) They do not display autosomal dominant inheritance.

(c) They were also not included in the HCM prevalence and penetrance studies that we will introduce later.

As a result, we ignore them in our study.

2.3.2 The Features of Genetic Testing in HCM

Since we aim to model the impact of genetic testing in HCM with an application to the life insurance market, we need to discuss the methodology and uptake of genetic testing in HCM. We do this in detail in Chapter 4.

2.4 HCM-related Endpoints

2.4.1 Risk, Features, and Management

Maron et al. (2014) describe three major causes of HCM-related mortality, affecting both genders to the same degree, associated with HCM-related endpoints.

(a) *Sudden cardiac arrest* (SCA) is often associated with no or mild symptoms (NYHA Class I/II in Table 2.1). This is the manifestation of HCM that

sometimes strikes healthy athletes. It is most prevalent among persons less than 30 years old. The annual rate of this event for persons at ages below or equal 20 was estimated at much higher rates compared to those of ages above 20 in Maron et al. (1999) and Spirito et al. (2000). It can be prevented with implantable cardioverter-defibrillator (ICD) treatment that can be offered when any major risk factor, such as increased LVWT, age, family history of SCA, syncope, or ECG abnormalities, have arisen.

(b) Death from *heart failure* (HF) is often associated with a history of severe symptoms (NYHA Class III/IV in Table 2.1) (Spirito et al. 2000, Maron et al. 2015, 2016a). It is not confined to any particular age range. It can be alleviated with invasive treatments if pharmacological treatment is not useful. Septal myectomy (alcohol ablation) or heart transplant are indicated depending on the particular clinical condition that presents in addition to increased LVWT.

(c) *Stroke death* is often caused by atrial fibrillation (AF). It is most often observed at older ages, and will not be significant for our purposes.

2.4.2 A Comment on Sudden Cardiac Death

We use the term 'sudden cardiac arrest' (SCA) and note that it may be fatal or non-fatal. This differs from the term used throughout the epidemiological literature, which is 'sudden cardiac death' (SCD). Surprisingly, at least to actuaries, this can also be fatal or non-fatal. A typical definition is that given in Elliott et al. (2006) (emphasis added):

"The following endpoints were used in the survival analysis: (1) sudden cardiac death — witnessed sudden death with or without documented ventricular fibrillation, death within one hour of new symptoms, nocturnal death with no antecedent history of worsening symptoms, and *successfully resuscitated cardiac arrest*; (2) ..."

The risk of confusion is obvious. For example, a review study determining the risk factors of HCM for health and life insurance applications by Pokorski (2002)

relied on the HCM-related mortality rates of a survival analysis published by Maron et al. (2000) which included non-fatal cases of SCD.

2.4.3 Relationship with Common Gene Mutations

Significant differences in the endpoints associated with different HCM-related mutations have not been established conclusively.

Watkins et al. (1995) found mutations in the MYH7 and TNNT2 genes to be associated with high SCA risk at ages less than 30 years. Niimura et al. (1998) found MYBPC3 mutation carriers to have a better prognosis compared to previous studies of MYH7 and TNNT2 mutation carriers. However, Van Driest et al. (2004) did not find MYBPC3 mutation carriers to have a more favourable prognosis than MYH7 mutation carriers.

Page et al. (2012) were not able to find a relationship between MYBPC3 gene mutations and the various clinical manifestations of HCM.

Given this uncertainty, we do not attempt, in our model, to distinguish between the endpoint after onset associated with mutations in different HCM-related genes, or different mutations in the same gene. We do distinguish between early-onset and late-onset HCM, but only before onset occurs; after onset we assume the same rates for the endpoints.

2.4.4 The Historical Pattern of HCM-Related Annual Mortality Rates

Here we consider the measurement of annual mortality rates, or hazard rates, associated with HCM-related endpoints.

(a) The earliest epidemiological studies of HCM tended to suggest extremely high mortality. However these studies were relatively small, and based on highly-selected populations, for example persons undergoing clinical treatment for symptoms of HCM. For example, Teare (1958) recorded seven fatal SCA endpoints among eight individuals with HCM at ages between 14 and 44 years. Such results are affected by *referral bias* or *patient selection bias*, (Maron

et al. 1999), and would not be applicable to a population of otherwise healthy carriers of an HCM-related mutation.

(b) Since Teare (1958) and until the 1990s (Maron et al. 1999), most epidemiological studies estimated HCM-related annual hazard rates as between 3% and 6%. Referral bias was still present, because the subjects were mostly severely symptomatic HCM patients who had been referred to tertiary hospitals rather than patients having a better clinical profile in the general population. The latter paper, a US regional study including 227 HCM patients at ages 1 month to 86 years, estimated an HCM-related annual hazard rate of 1.3% if referral bias was absent.

(c) An important feature of any epidemiological study is the definition of the endpoint. In an actuarial mortality study this is always death. In a clinical study, this is not necessarily so. We noted in Section 2.4.2 that the endpoint often used in epidemiological studies — sudden cardiac death — is not always fatal. Studies generally do report fatal and non-fatal endpoints separately, but often combine them in the published survival analysis. We give two examples below.

(i) Maron et al. (2000) was a cohort study of 744 patients over a wide age range, in Italy and the USA. Non-fatal SCD and heart transplant were recorded as HCM-related fatal endpoints. The HCM-related annual hazard rate was estimated to be 1.4%. However, the annual hazard rate would drop to 1.08% if non-fatal endpoints were excluded.

(ii) Elliott et al. (2006) was a cohort study of 956 individuals affected by HCM at ages 16 to 88 years in the UK. The annual hazard rate of SCD was estimated to be 1%, but would drop to 0.4–0.8% if non-fatal SCD was excluded. Table 3.4 estimates the impact of recalculating the annual hazard rates reported in several studies if non-fatal endpoints were excluded.

(iii) Modern diagnosis and treatment techniques have brought HCM-related

annual hazard rates below 1%, in population studies in Europe and the USA (Elliott et al. 2006). In very recent studies, HCM-related annual hazard rates were reported to be as low as 0.5% for all ages (Maron et al. 2013, 2015, 2016a). These low mortality rates were explained by ICD treatment to avert fatal SCA and surgery to avert heart failure.

In what follows, we rely on the three very recent studies Maron et al. (2013), Maron et al. (2015) and Maron et al. (2016a). These are relatively large, avoid referral bias as far as possible and distinguish between fatal and non-fatal HCM-related endpoints.

2.5 Prevalence of HCM

Prevalence estimates of HCM-related mutations in the general population depend on two types of studies, one of which was based on clinically affected individuals, the other on the analysis of DNA.

2.5.1 Prevalence of Clinical HCM

Maron et al. (1995) examined the prevalence of clinical HCM in the general population. The study included 4,111 men and women at ages 25 to 35 years in the USA, of whom seven were diagnosed with HCM. (This sample was obtained by a random selection in the general population, which means these individuals were clinically unknown, or undetected.) The estimated prevalence was approximately 0.2%. Other studies of HCM in the general population, in Japan, China, and East Africa, estimated similar prevalence rates (Maron 2004, Semsarian et al. 2015).

The recent studies of clinical prevalence of HCM (Maron et al. (2016b), Husser et al. (2018)) presented the results of 'claims-based analysis', the analysis of very large healthcare databases in which people were recorded on clinically diagnosed with HCM.

(a) In Maron et al. (2016b), a medical database with 169,098,614 patients (more than half of the US population) was examined, of whom, 59,009 HCM-related

claims were found, a prevalence of approximately 0.035%.

(b) In Husser et al. (2018), a German healthcare claims database with 5,490,810 patients was examined, of whom, 4,000 cases were found, a prevalence of approximately 0.07%.

That is, their findings show that the prevalence of clinical HCM in the general population varies between 0.035–0.07%. This has interesting results.

(I) Onset of HCM does not necessarily refer to diagnosis of HCM since the majority of individuals who developed HCM are thought to be 'silent'.

(II) HCM is clinically diagnosed (or detected) in several ways: clinical diagnosis of asymptomatic patients based on family history, routine checks or incidental findings, or based on symptoms (the majority of cases in this subject) before or after an HCM-related event such as non-fatal or fatal HCM. See Maron et al. (1982), Adabag et al. (2006), and Elliott et al. (2006).

2.5.2 Prevalence of HCM-Related Mutations

Bick et al. (2012) studied the prevalence of HCM-related mutations in the general population. The study included 3,600 men and women at ages 30 to 84 years, of whom twenty-two carried known HCM-related mutations. The estimated prevalence rate of mutations was 0.6%, higher than the level estimated by studies of clinical HCM. Moreover, only four of these twenty-two genetically affected individuals had clinical HCM.

The difference between these results has interesting consequences.

(a) As we noted in Section 2.3.1, not all HCM-related mutations have been identified. Bick et al. (2012) could detect only those HCM-related mutations that were known at the time, so 0.6% is a minimum, the prevalence of mutations might be as high as 1%.

(b) The study also shows that not everyone who is a carrier of an HCM mutation will ever develop the disorder.

(c) If there is a much larger pool of mutation carriers than clinical observations of HCM suggest, then rates of onset and mortality among mutation carriers may be much lower than previously thought. If genetic test results were to encourage mutation carriers to over-insure (adverse selection) then for every such new insurance contract that brings the clinical risk of HCM observed in Maron et al. (1995) to the insurance pool, there may be another two that do not.

2.6 From the Epidemiology of HCM to a Mathematical Model of HCM

In this chapter we have given a broad outline of the epidemiology of HCM. Before we go into any more detail, it is convenient to consider how we might formulate a mathematical model of HCM that will allow us to address actuarial questions. We do this in Chapter 3. The model we propose is a multiple-state Markov model, whose key parameters are the hazard rates or transition intensities between the states. To fully specify the model we then need to estimate these hazard rates. To do this we return to a more detailed examination of the epidemiology of HCM, in Sections 3.7, 3.8, and 3.9.

Chapter 3

A Mathematical Model of the Epidemiology of Hypertrophic Cardiomyopathy (HCM)

3.1 Introduction

In this chapter, we model the epidemiology of HCM using a multiple-state Markov model. For simplicity, at the moment, the model does not include genetic testing because the nature of genetic testing in HCM cannot be easily modelled in a Markovian setting. Chapter 4 describes genetic testing in HCM and Chapter 5 represents it in a multiple-state model.

In this chapter, we model only HCM-related events such as onset of HCM, non-fatal or fatal HCM. This model will be referred to as 'the epidemiological model' (of HCM). In Section 3.2, we introduce a mathematical model of the epidemiology of HCM. In Section 3.3, we formulate the epidemiological model of HCM and express the occupancy probabilities in the model states in terms of the Kolmogorov forward equations. Then, in Sections 3.4 and 3.5, we derive and numerically solve the Kolmogorov forward equations. In Section 3.6, we survey the necessary parameters (transition intensities) of the model. In Sections 3.7, 3.8, 3.9, and 3.10, we estimate these model parameters. In Section 3.11, we show examples of occupancy probabilities in the model states.

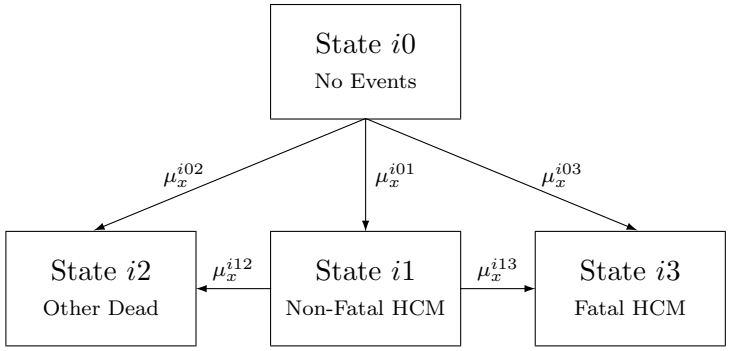


Figure 3.1: A mathematical model of HCM, representing the life history of a person in the ith of several sub-populations defined by HCM genotype.

3.2 The Epidemiological Model of HCM

In what follows we denote age by x and time by $s \in [0, \infty]$ after age x. Here we only model a life history of a single individual, however, in later chapters, it will be necessary to describe age x as a function of 'calendar time' $t \in [0, \infty]$.

Our mathematical model of HCM is the discrete-state continuous-time Markov model shown in Figure 3.1.

Assumption 3.1. In the model, a foundational assumption is the Markov assumption itself which is that the probability of being in any model state in future at age $x + s$ is only conditional on the currently occupied state at age x and not any other past history (see equation (3.2)).

It would be reasonable to suppose that transition intensities can be influenced by previous events in a life history. However, it is generally too difficult to specify what these might be and how they affect the intensities, so we choose a Markov model for simplicity. Where it is essential to condition intensities on an event other than age and the currently occupied state (as will be the case in Chapter 5 when we model genetic testing in HCM) then we must use a non-Markov model at the expense of more complex computations.

The entire population is represented by a collection of five such models, see Figure 3.2:

(a) four sub-populations representing those who carry an HCM-related mutation

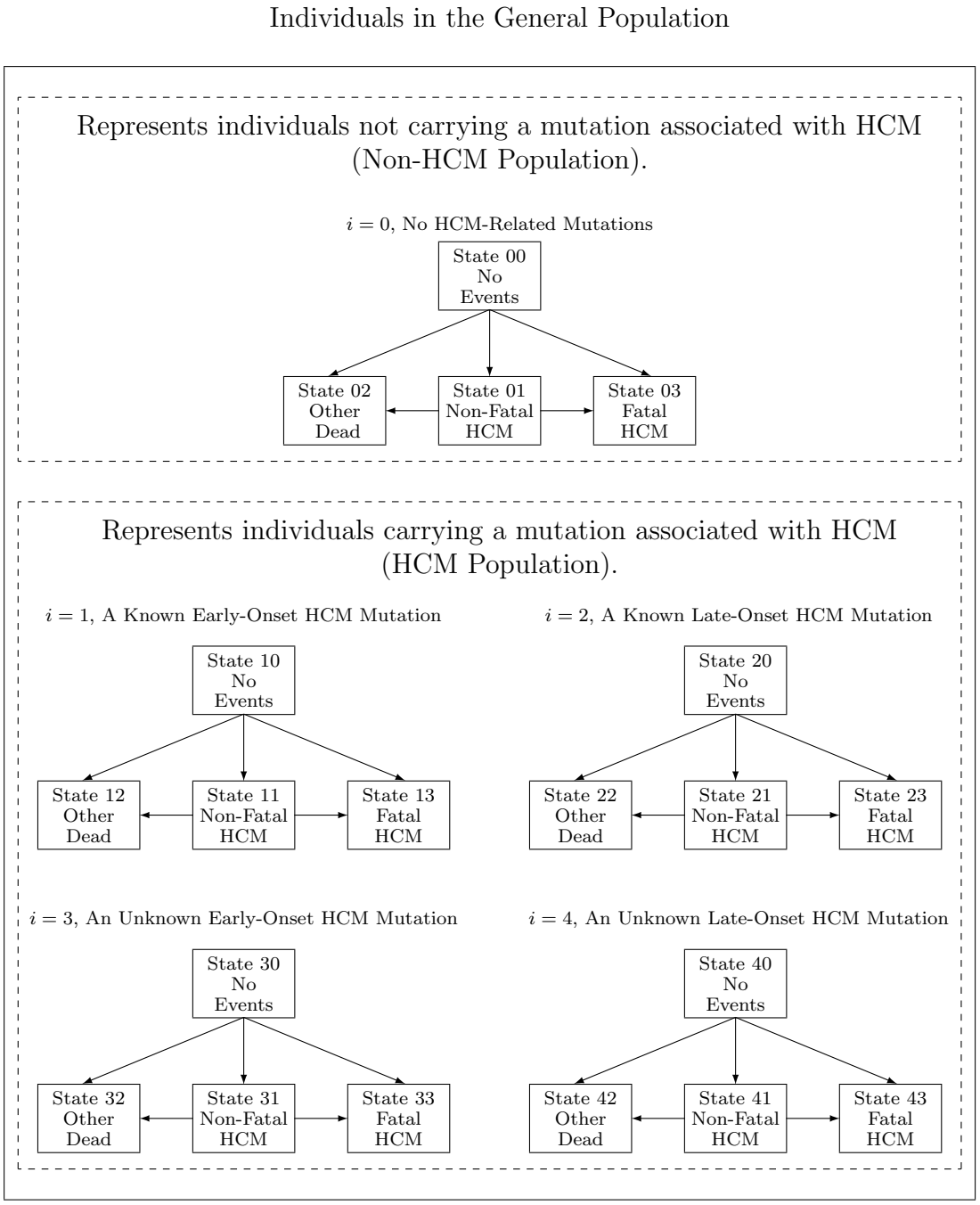


Figure 3.2: A mathematical model of HCM for a population consists of non-HCM population with sub-population $i = 0$, carrying no HCM-related mutations and HCM population with the collection of sub-populations $i = 1$, carrying a known early-onset HCM mutation; $i = 2$, carrying a known late-onset HCM mutation; $i = 3$, carrying an unknown early-onset HCM mutation; and $i = 4$, carrying an unknown late-onset HCM mutation.

based on genotype defined in Section 2.3.1, in which the collection of these sub-populations is referred to as the 'HCM population'; and

(b) one sub-population representing those who do not carry any HCM-related mutation which is referred to as the 'non-HCM population'.

Sub-populations are labelled by 'i' as follows:

- $i = 0$: individuals who do not carry any HCM-related mutations.
- $i = 1$: individuals who carry a known early-onset HCM-related mutation.
- $i = 2$: individuals who carry a known late-onset HCM-related mutation.
- $i = 3$: individuals who carry an unknown early-onset HCM-related mutation.
- $i = 4$: individuals who carry an unknown late-onset HCM-related mutation.

Thus, we can represent different risks associated with different HCM-related mutations by different transition intensities in the models representing the respective sub-populations.

3.3 The Formulation of the Model

3.3.1 The Probabilities at Birth

We assume that a person chosen at random occupies at birth, one of the states 00, 10, 20, 30, or 40, with respective probabilities equal to the population prevalence of genotype i at birth:

(a) p_{00} is the prevalence of non-carriers of HCM-related mutations at birth,

(b) p_{10} is the prevalence of known early-onset mutation carriers at birth,

(c) p_{20} is the prevalence of known late-onset mutation carriers at birth,

(d) p_{30} is the prevalence of unknown early-onset mutation carriers at birth,

(e) p_{40} is the prevalence of unknown late-onset mutation carriers at birth,

and

$$\sum_{i=0}^{4} p_{i0} = 1. \tag{3.1}$$

3.3.2 Transition Intensities and Occupancy Probabilities after Birth

After birth, the epidemiological model (see Figures 3.1 and 3.2) is parametrized by transition intensities labelled by μ_x^{ijk}. They are instantaneous rates of transitions between model states which are defined in terms of the occupancy probabilities,

$$ {}_sp_x^{ijk} = P[\text{In state } ik \text{ at age } x+s \mid \text{In state } ij \text{ at age } x], \tag{3.2} $$

as follows:

$$ \mu_x^{ijk} = \lim_{ds \to 0} \frac{{}_{ds}p_x^{ijk}}{ds}, \quad j \neq k. \tag{3.3} $$

Assumption 3.2. In the other direction, the occupancy probabilities can be obtained from transition intensities as follows:

$$ {}_{ds}p_x^{ijk} = \mu_x^{ijk} ds + o(ds), \quad j \neq k. \tag{3.4} $$

Assumption 3.3. We assume that the probability of two or more transitions in small time ds is $o(ds)$.

Then, at any age $x \geq 0$, we have:

$$ \sum_{i=0}^{4} p_{i0} \sum_{k=0}^{3} {}_sp_0^{i0k} = 1. \tag{3.5} $$

That is, at birth no-one has suffered onset of HCM, and at any later age x the law of total probability holds.

We obtain the occupancy probabilities from transition intensities by solving the Kolmogorov forward equations in Section 3.4. As usual, we will parametrise the model by estimating the transition intensities and we will use the Kolmogorov forward equations to find occupancy probabilities when needed.

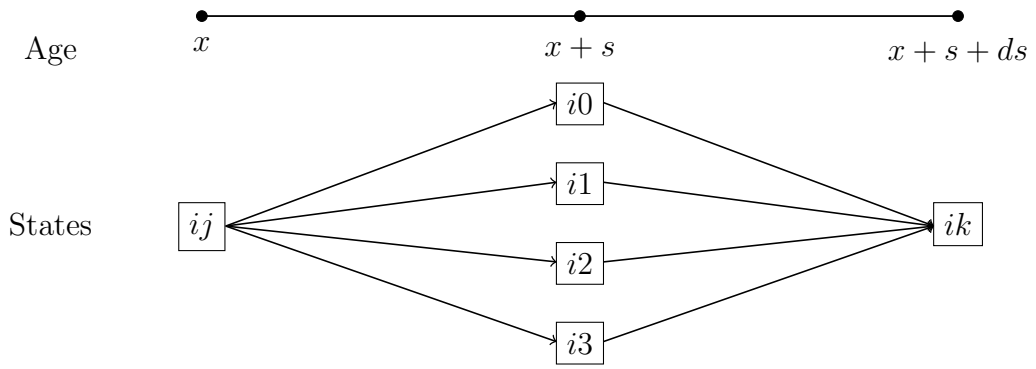


Figure 3.3: Generalized form of all probabilistic paths of a person in state ij at age x into state ik at age $x+s+ds$.

3.3.3 Transition Intensities and Clinical HCM

An HCM-related event can occur only if the carrier of an HCM-related mutation has suffered onset of HCM (see Section 2.2.2) whether or not symptoms are present. We do not model onset of HCM as a transition between states. Instead, we define the penetrance of clinical HCM, denoted by $F(x)$, as:

$$F(x) = P[\text{Onset of HCM has occured by age } x]. \tag{3.6}$$

If the hazard rates of non-fatal and fatal HCM-related events, conditional on onset having occurred, are denoted by ρ_x^{i01} and ρ_x^{i03} respectively, then the hazard rates assuming onset to be unobserved are $\mu_x^{i01} = F(x)\rho_x^{i01}$ and $\mu_x^{i03} = F(x)\rho_x^{i03}$. These are the hazard rates used in the model. Other transition intensities are as indicated in Figure 3.1.

3.4 The Kolmogorov Forward Equations

Here we will derive the Kolmogorov forward equations (Macdonald et al. (2018) is a good reference for this purpose). They form a system of ordinary differential equations (ODEs).

We derive the general form of the Kolmogorov forward equations in five steps:

1. We denote by Ψ the total number of states in the model. The model has twenty states, see Figure 3.2, therefore, $\Psi = 20$ at the moment.

2. We show the general form of possible transitions between the model states after age x in our model in Figure 3.3.

3. Then, we can write down the occupancy probability of a person in state ik at age $x+s+ds$ given that the person was in state ij at age x in our model as follows:

$$ {}_{s+ds}p_x^{ijk} = {}_sp_x^{ijk}\,{}_{ds}p_{x+s}^{ikk} + \sum_{\substack{l=0\\l\neq k}}^{\Psi} {}_sp_x^{ijl}\,{}_{ds}p_{x+s}^{ilk}. \tag{3.7} $$

4. Since the total probability of being in all model states at any specific time is always equal to 1, we can express ${}_{ds}p_{x+s}^{ikk}$, see assumptions 3.2 and 3.3, as follows:

$$ {}_{ds}p_{x+s}^{ikk} = \left(1 - \sum_{\substack{l=0\\l\neq k}}^{\Psi} \left(\mu_{x+s}^{ikl}ds + o(ds)\right)\right) \tag{3.8} $$

5. Therefore, equation (3.7) is represented by:

$$ \begin{aligned} {}_{s+ds}p_x^{ijk} = {}_sp_x^{ijk} &\left(1 - \sum_{\substack{l=0\\l\neq k}}^{\Psi} \left(\mu_{x+s}^{ikl}ds + o(ds)\right)\right) \\ &+ \sum_{\substack{l=0\\l\neq k}}^{\Psi} {}_sp_x^{ijl}\left(\mu_{x+s}^{ilk}ds + o(ds)\right) \end{aligned} \tag{3.9} $$

Now, subtract ${}_sp_x^{ijk}$ from both sides of the equation,

$$ {}_{s+ds}p_x^{ijk} - {}_sp_x^{ijk} = \sum_{\substack{l=0\\l\neq k}}^{\Psi} {}_sp_x^{ijl}\left(\mu_{x+s}^{ilk}ds + o(ds)\right) - {}_sp_x^{ijk}\sum_{\substack{l=0\\l\neq k}}^{\Psi}\left(\mu_{x+s}^{ikl}ds + o(ds)\right), \tag{3.10} $$

divide the each side of the equation by ds and take the limit $ds \to 0^+$, where $\lim_{ds\to 0^+}\left(o(ds)/ds\right) = 0$,

$$ \lim_{ds\to 0^+} \frac{{}_{s+ds}p_x^{ijk} - {}_sp_x^{ijk}}{ds} = \sum_{\substack{l=0\\l\neq k}}^{\Psi} {}_sp_x^{ijl}\mu_{x+s}^{ilk} - {}_sp_x^{ijk}\sum_{\substack{l=0\\l\neq k}}^{\Psi}\mu_{x+s}^{ikl}, \tag{3.11} $$

and obtain the general form of the Kolmogorov forward equations as follows:

$$\frac{d}{ds}{}_sp_x^{ijk} = \underbrace{\sum_{\substack{l=0\\l\neq k}}^{\Psi} {}_sp_x^{ijl}\mu_{x+s}^{ilk}}_{\substack{\text{rate of entering}\\ \text{into state } ik \text{ at}\\ \text{age } x+s}} - \underbrace{{}_sp_x^{ijk}\sum_{\substack{l=0\\l\neq k}}^{\Psi}\mu_{x+s}^{ikl}}_{\substack{\text{rate of leaving}\\ \text{from state } ik \text{ at}\\ \text{age } x+s}}. \tag{3.12}$$

3.5 Numerical Solution of the Kolmogorov Forward Equations

Several numerical methods can be applied to solve the Kolmogorov forward equations. Macdonald et al. (2018) discusses two methods: *the Euler method* as the simplest; and *the fourth-order Runge-Kutta method* as being much more efficient. Thus we follow the latter. Firstly, formulate any component of equation (3.12) as

$$d\left({}_sp_x^{ijk}\right) = f\left(s, {}_sp_x^{ijk}\right) ds. \tag{3.13}$$

Secondly, apply the algorithm of *the fourth-order Runge-Kutta method* which solves the equations as follows:

$${}_{s+ds}p_x^{ijk} \approx {}_sp_x^{ijk} + \frac{dp^1 + 2dp^2 + 2dp^3 + dp^4}{6} \tag{3.14}$$

where

$$dp^1 = f(s, {}_sp_x^{ijk})ds, \tag{3.15}$$

$$dp^2 = f\left(s + ds/2, {}_sp_x^{ijk} + dp^1/2\right) ds, \tag{3.16}$$

$$dp^3 = f\left(s + ds/2, {}_sp_x^{ijk} + dp^2/2\right) ds, \tag{3.17}$$

$$dp^4 = f\left(s + ds, {}_sp_x^{ijk} + dp^3\right) ds. \tag{3.18}$$

We solve these equations forward ($ds > 0$), since the boundary conditions for the occupancy probabilities are ${}_0p_0^{i00} = p_{i0}$, the prevalence rate at birth in the ith subpopulation and ${}_0p_0^{i0k} = 0$ where $k \neq 0$. Now, we can find the occupancy probability

Table 3.1: The prevalence rate of HCM in the general population. LVWT: Left Ventricular Wall Thickness.

Population	Diagnosis Baseline (Criteria)	Age (yr)	Prevalence Rate	Reference
The study included unrelated 4,111 men and women in the USA, of whom seven were *clinically* diagnosed with HCM.	Clinical (LVWT≥15 mm)	23–35	0.2%	Maron et al. (1995)
The study included unrelated 3600 adults men and women in the USA, of whom twenty-two carried known HCM-related mutations.	Genetics (HCM causing known gene mutations)	30–84	0.6%	Bick et al. (2012)

at any age in any model state.

3.6 The Parameters of the Model

To fully parametrise this model, therefore, we need estimates of the following.

(a) The population prevalences p_{i0} at birth.

(b) The age-related onset of clinical HCM, see $F(x)$ (equation (3.6)) in Section 3.3.3, in the mutation-carrying sub-populations.

(c) Hazard rates of fatal and non-fatal HCM-related events after onset of clinical HCM. Referring to Section 2.4.3, we assume the same hazard rates after onset of HCM for the carriers of different HCM-related mutations.

(d) Mortality rates from all non-HCM-related causes. We assume these to be the same in all five sub-populations.

Ages zero to 20 are influential, because we will assume that no life insurance is purchased before age 20. Onset and mortality rates are high, among early-onset mutation carriers. High rates of onset increase the number who reach age 20 with clinical HCM, while high rates of mortality decrease that number.

3.7 Parametrising the Model I: Prevalence

The prevalence of HCM-related mutations is here considered in two stages.

Table 3.2: The prevalence rate of HCM-related mutations in the HCM population conservatively estimated from the reported rates in Elliott et al. (2014).

	Mutation	Early-Onset	Late-Onset	Total
Baseline	Known	50%	16.67%	66.67%
	Unknown	25%	8.33%	33.33%
	Total	75%	25%	100%
Sensitivity	Known	56.25%	18.75%	75%
	Unknown	18.75%	6.25%	25%
	Total	75%	25%	100%

(a) We firstly consider the prevalence rate of HCM-related mutations in the general population. See all the model sub-populations in Figure 3.2.

(b) We secondly consider the prevalence rates of HCM-related mutations in the different sub-populations (see Sections 3.2 and 3.3.1) in the HCM population. See the model sub-populations $i = 1$, $i = 2$, $i = 3$, and $i = 4$ in Figure 3.2.

3.7.1 Prevalence of HCM Mutations: General Population

We follow two reported prevalence studies of HCM in the general population, summarised in Table 3.1 (see Section 2.5).

(a) Maron et al. (1995) reports that the prevalence of clinical HCM is about 0.2% in the general population at ages 23–35 years. See Section 2.5.1.

(b) Bick et al. (2012) reports that the prevalence of HCM-related known mutations is 0.6% in the general population at ages 30–84 years. See Section 2.5.2.

We choose to use a prevalence of 0.2% as our baseline for the prevalence rate of HCM-related mutations in the general population because it is conservative. It assumes the prevalence of HCM-related mutations and clinical HCM to be the same. Later, in Chapter 9, we investigate the consequences of a much higher prevalence of HCM-related mutations such as the prevalence figure in Bick et al. (2012) above.

3.7.2 Prevalence of HCM Mutations: HCM Population

Table 3.2 shows, from Elliott et al. (2014), the prevalence rates of HCM mutations in the HCM population, which are conservative for our purposes. See Section 2.3.1.

(a) We assume, ignoring irrelevant mutations, the prevalence rate of known mutations in the HCM population is about 2/3 (baseline) to 3/4 (sensitivity).

(b) We assume 3/4 of known mutations (and the same for unknown mutations) are early-onset mutations.

(c) Elliott et al. (2014) does not report a specific age for these prevalences. For simplicity, they are assumed to be at birth (age zero) in this study.

3.8 Parametrising the Model II: Onset

The *penetrance* of a mutation is the probability that the phenotype associated with the mutation is actually present. See Section 3.3.3. If penetrance is less than 100% it is said to be incomplete. Not many late-onset single-gene disorders of humans have complete penetrance; it is the exception rather than the rule. If the phenotype may appear sometime after birth, then we can define the *age-related* penetrance of a specific mutation, denoted by the non-decreasing function $F(x)$, see equation (3.6). If penetrance is incomplete then $\lim_{x\to\infty} F(x) < 1$. Except for the last property, $F(x)$ behaves in every way like a cumulative distribution function.

The penetrance of HCM has been estimated to be 69% in Charron et al. (1997), which was regarded as the baseline penetrance of HCM in Howard (2014), see Section 10.2. However, this must be qualified.

(a) Clinical expression of HCM is heteregenous. In HCM, mutations in different genes, and different mutations in the same gene, may have different penetrances.

(b) If there is a high proportion of 'silent' mutations, see Section 2.5, penetrance estimates relying on clinically diagnosed HCM may be overstated. This is discussed in Section 9.3.1 and its impact in adverse selection costs is significant.

3.8.1 Penetrance of Early-Onset HCM Mutations

Penetrance estimates of the form of equation (3.6) are not available below age 20. This is partly related to the difficulties in carrying out medical studies involving young children, and partly the uncertainty associated with clinically defining onset at these ages (see Section 2.2.2). However we may rely on the general observations that onset often occurs in adolescence and early adulthood; and mortality is at its highest up to about age 30 (see Section 3.9).

For our purposes, a conservative assumption is that early-onset mutations have 100% penetrance by age 20. We are unable to estimate directly the proportion of mutation carriers who develop HCM and die before age 20. Yet, in our study, the penetrance before age 20 is not particularly relevant. So, we gradually increase the early-onset penetrance up to age 20 by assuming (same for both genders) $F(x) = 0.25$ below age 10; $F(x) = 0.5$ at ages 10 to below 15; and $F(x) = 1$, afterwards.

3.8.2 Penetrance of Late-Onset HCM Mutations

Late-onset HCM is mainly associated with mutations in the MYBPC3 gene (see Section 2.3.1.1). Thus a mutation carrier may be identified as being at risk by a genetic test alone, when no clinical symptoms are present. Two studies have estimated the age-related penetrance of such mutations:

(a) Christiaans et al. (2011) studied 446 mutation carriers in 166 families in the Netherlands (44% male, 56% female). The majority had the same mutation in the MYBPC3 gene, explained by there being founders of Dutch origin in the past. None had been clinically evaluated as having HCM. Figure 3.4 shows their Kaplan-Meier estimates of penetrance by age x for males and females, as

well as our own smoothed estimates:

$$F(x)^{\text{Dutch Males}} = 5.049(10^{-3}) - 9.488(10^{-4})x + 9.550(10^{-5})x^2 + 2.636(10^{-6})x^3 - 2.428(10^{-8})x^4 \quad (3.19)$$

$$F(x)^{\text{Dutch Females}} = -3.743(10^{-3}) + 3.868(10^{-3})x - 3.319(10^{-4})x^2 + 9.348(10^{-6})x^3 - 4.991(10^{-8})x^4. \quad (3.20)$$

The authors noted that their penetrance estimates were lower than in other studies because they excluded known clinically affected probands (the index patient in a family affected by HCM) and relatives.

(b) Terauchi et al. (2015) studied 61 MYBPC3 mutation carriers in 28 families in Japan (%51 male, %49 female). Their penetrance estimates showed a similar pattern by age and gender to those of Christiaans et al. (2011) but were somewhat higher. Figure 3.5 shows their Kaplan-Meier estimates of penetrance by age x for both gender, as well as our own smoothed estimates:

$$F(x)^{\text{Japan Males}} = 1.608(10^{-2}) - 6.861(10^{-3})x + 2.040(10^{-4})x^2 + 9.051(10^{-6})x^3 - 1.120(10^{-7})x^4 \quad (3.21)$$

$$F(x)^{\text{Japan Females}} = 1.806(10^{-3}) + 2.306(10^{-3})x - 7.659(10^{-5})x^2 + 5.757(10^{-6})x^3 - 4.066(10^{-8})x^4. \quad (3.22)$$

We use Christiaans et al. (2011) as our baseline. (Note that sensitivity testing, in respect of adverse selection costs, in Section 9.3.2 shows that the choice of either study is unimportant.) We assume $F(x) = 0$ below age 20, $F(x)$ is given by equations (3.19) and (3.20) for males and females respectively at ages 20–70, and $F(x) = 1$ above age 70. For sensitivity analysis, we use Terauchi et al. (2015), substituting equations (3.21) and (3.22) for equations (3.19) and (3.20).

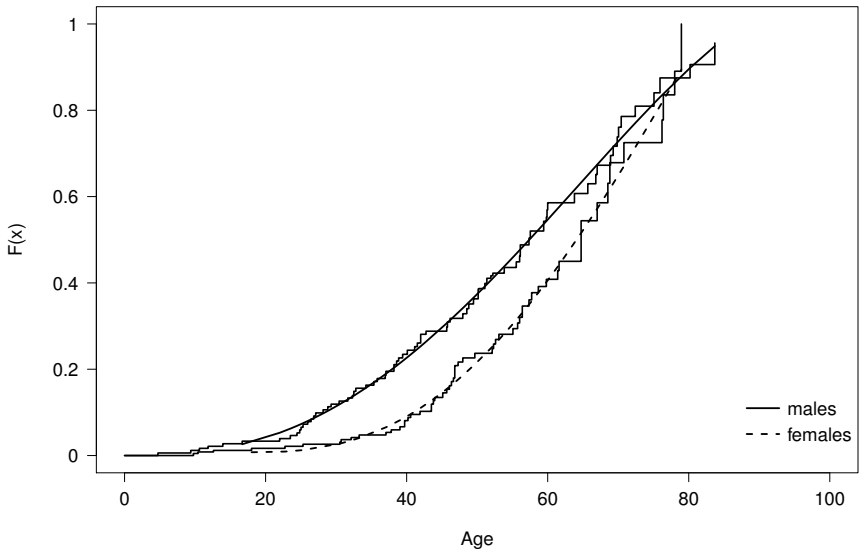


Figure 3.4: Late-Onset Penetrance of HCM from Christiaans et al. (2011)

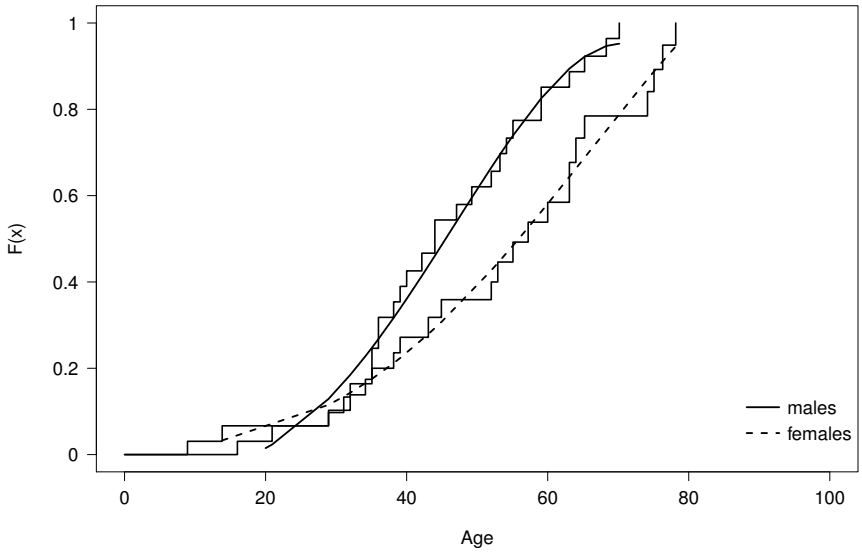


Figure 3.5: Late-Onset Penetrance of HCM from Terauchi et al. (2015)

3.9 Parametrising the Model III: HCM-Related Endpoints

3.9.1 Data Sources

In Section 2.4.4, we noted that estimates of the mortality rates associated with HCM had fallen steadily since first being studied, as a result of referral bias being reduced, and methods of diagnosis and treatment being improved.

We also noted, in Section 2.4.2, the need to distinguish between fatal and non-fatal events among the endpoints of any study. Therefore, we make the following definitions.

(a) *Fatal HCM endpoints:* sudden cardiac arrest causing death (Fatal SCA), death from progressive heart failure (HF), stroke, and post-operative deaths.

(b) *Non-fatal HCM endpoints:* resuscitated cardiac arrest and heart transplant.

For these reasons, we base our estimates of HCM-related hazard rates on three large, recent studies. Maron et al. (2016a, 2015, 2013) observed 1,902 individuals clinically affected by HCM, from the Minneapolis Heart Institute and Tufts Medical Center in the USA. The ages at onset of HCM of these individuals were not known; instead each was observed from an initial evaluation age, at which clinical HCM was established. These ages ranged from seven to 91 years old. Thus, observation of HCM-related endpoints is left-truncated (see Macdonald et al. (2018)). Each of the three studies analysed a different range of initial evaluation ages.

- Maron et al. (2016a) included 474 individuals with initial evaluation ages seven to 29 years old, subdivided into age ranges 7–10, 11–15, 16–20, 21–25, and 26–29. Three deaths from heart transplant complications were recorded as HF deaths.

- Maron et al. (2015) included 1,000 individuals with initial evaluation ages 30 to 59 years old, subdivided into age ranges 30–39, 40–49, and 50–59. Five deaths from heart transplant complications were recorded as HF deaths, and

Table 3.3: HCM-related endpoints from Maron et al. (2016a, 2015, 2013). The fatal endpoints are Fatal SCA: Sudden Cardiac Arrest causing death, HF: Heart Failure, Stroke and P-O: Post-Operative. The non-fatal endpoints are ReCA: Resuscitated Cardiac Arrest, and HT: Heart Transplant.

Initial Evaluation Ages	No.	Fatal Endpoints: Fatal	SCA	HF	Stroke	P-O	Non-Fatal Endpoints: ReCA	HT	Reference
7–29	474	12	5	0	1	20	12	Maron et al. (2016a)
7–10	10	0	0	0	0	1	0	
11–15	94	5	0	0	0	1	2	
16–20	173	5	4	0	0	10	3	
21–25	117	0	0	0	1	5	4	
26–29	80	2	1	0	0	3	3	
30–59	1000	17	17	2	4	5	20	Maron et al. (2015)
30–39	290	7	5	1	0	3	8	
40–49	361	6	7	0	3	1	8	
50–59	349	4	5	1	1	1	4	
60–91	428	2	2	6	2	3	1	Maron et al. (2013)

two heart transplants following resuscitated cardiac arrest were recorded as heart transplants.

- Maron et al. (2013) included 428 individuals with initial evaluation ages 60 to 91 years.

For brevity, we will refer to these three papers in what follows as 'the Maron et al. papers'. In Table 3.3, we summarise the results of these three studies.

3.9.2 HCM-Related Hazard Rates

In our subsequent modelling of term insurance contracts, we will focus on ages 20 to 60, as being a representative range of ages when such contracts are in force. We have little interest in ages before 20, except insofar as deaths before age 20 reduce the population who might buy insurance after age 20. Our aim here is to use the results in the Maron et al. papers to estimate the hazard rates of HCM-related mortality between ages 20 and 60.

3.9.3 The Estimation of HCM-Related Hazard Rates

If we observe a group of n_x persons with age label x exposed to the risk of some event for a total time of E_x^c years, and record d_x events, we may model the number of events as a random variable D_x having a Poisson distribution with parameter $\rho_x E_x^c$, where ρ_x is the force or hazard rate of the event applicable at the ages labelled x (note that we denote the hazard rates of non-fatal HCM and fatal HCM at age x by ρ_x in Section 3.3.3), as follows:

$$P(D_x = d_x) = \frac{e^{-\rho_x E_x^c}(\rho_x E_x^c)^{d_x}}{d_x!}. \tag{3.23}$$

Standard results lead to the maximum likelihood estimate (Macdonald et al. (2018)):

$$\hat{\rho}_x = d_x / E_x^c \tag{3.24}$$

of the hazard rate ρ_x, and its approximate sampling variance:

$$v_x \approx \hat{\rho}_x / E_x^c. \tag{3.25}$$

The age label x may refer either to a single age or a range of ages, over which the hazard rate is assumed to be constant.

The event of interest may be death, or other endpoints from different causes such as non-fatal HCM, taking d_x to be the number of relevant events observed.

3.9.4 Confidence Intervals of HCM-Related Hazard Rates

Approximate 95% confidence intervals for ρ_x are given by $\hat{\rho}_x \pm 1.96(\hat{\rho}_x / E_x^c)^{1/2}$. In some cases, this results in confidence intervals including negative values, which suggests that the usual asymptotic Normal theory on which these intervals are based is a poor approximation for these data. A better approach may be to estimate confidence intervals by parametric bootstrapping.

For each age label x, 10,000 values of a Poisson random variable with parameter $\hat{\rho}_x E_x^c$ are simulated, using R. This gives a vector of simulated numbers of events

consistent with the Poisson assumption, denoted by $\tilde{\boldsymbol{D}}_{\boldsymbol{x}}$:

$$\tilde{\boldsymbol{D}}_{\boldsymbol{x}} = (\tilde{D}_x^1, \tilde{D}_x^2, ..., \tilde{D}_x^{10000}), \tag{3.26}$$

Dividing each of these by E_x^c, we obtain a simulated sample of estimated hazard rates consistent with the Poisson assumption, denoted by $\tilde{\boldsymbol{\rho}}_{\boldsymbol{x}}$:

$$\tilde{\boldsymbol{\rho}}_{\boldsymbol{x}} = \frac{\tilde{\boldsymbol{D}}_{\boldsymbol{x}}}{E_x^c} = (\tilde{\rho}_x^1, \tilde{\rho}_x^2, ..., \tilde{\rho}_x^{10000}). \tag{3.27}$$

Approximate 95% confidence intervals can be read off directly from the ordered list of elements of $\tilde{\boldsymbol{\rho}}_{\boldsymbol{x}}$.

3.9.5 The Central Exposure to Risk (E_x^c) of HCM-Related Events

The Maron et al. papers do not provide the central exposure to risk E_x^c for any age groups. They do, however, state the mean follow-up times observed in respect of the aggregated age range in each study. At ages 7–29, the mean follow-up time was 7.1 years; at ages 30–59, it was 7.2 years; and at ages 60 and over, it was 5.8 years. We can therefore use the identity:

$$\text{Central exposure to risk=Number of persons} \times \text{Mean follow-up time} \tag{3.28}$$

to find the central exposures to risk for these aggregated age groups. Mean follow-up times for the subdivided age ranges in Maron et al. (2016a) and Maron et al. (2015) were not stated.

In Table 3.4, we present age-specific HCM-related annualized hazard rates, based on the data in Table 3.3. We show hazard rates including and excluding the non-fatal endpoints of resuscitated cardiac arrest and heart transplant. We note that mortality due to stroke is relatively unimportant below age 60.

Table 3.4: Annualized hazard rates of SCA and heart failure, and the overall hazard rate of HCM, are shown both including (inc.) and excluding (exc.) non-fatal endpoints. For the abbreviations used in the headings, please see the caption of Table 3.3. Note that the numerics are rounded to six decimal places.

Ages	n_x	E_x^c	SCA		HF		Stroke	HCM	
			(exc. ReCA)	(inc. ReCA)	(exc. HT)	(inc. HT)		(exc. non-fatal)	(inc. non-fatal)
7–29	474	3365.4	0.003566	0.009509	0.001486	0.005051	0.000000	0.005349	0.014857
30–59	1000	7200	0.002361	0.003056	0.002361	0.005139	0.000278	0.005556	0.009028
60–91	428	2482.4	0.000806	0.002014	0.000806	0.001209	0.002417	0.004834	0.006445

3.9.6 Obtaining HCM-Related Hazard Rates for Ages 20 to 60

From the Maron et al. papers, we have estimated hazard rates for ages 7–29, 30–59 and 60 and over. We also have numbers of events, but not mean follow-up times, for subdivisions of the first two age ranges. Our purpose requires hazard rates for ages 20–60. We consider three alternative assumptions here how we may adjust the data we have, in order to approximate the required hazard rates.

Assumption 3.4. $\rho_{[20-29]} = \rho_{[7-29]}$: Assume hazard rates to be constant over the age ranges shown in Table 3.4, then use the hazard rate for ages 7–29 at ages 20–29. (We assume the estimated hazard rates for ages 30–59 apply up to exact age 60.)

Assumption 3.5. $\rho_{[20-29]} = \rho_{[30-59]}$: Assume hazard rates for ages 30–59 also apply to the age range 20–29. This assumption might be justified for fatal HCM since the numbers of HCM-related fatal events at ages 21–29 were very small (see Table 3.3).

Assumption 3.6. Assume the 7.1 years mean follow-up for ages 7–29 also applies at ages 21–29, so we can find the central exposure to risk E_x^c for the latter age range. Then combining the numbers of HCM-related events and the exposures for the age ranges 21–29 and 30–59, we can calculate a single constant hazard rate for ages 21–59. We then assume this hazard applies also in the year of age 20–21.

3.9.7 Estimated HCM-Related Hazard Rates

Table 3.5 shows the estimated hazard rates of SCA and heart failure, and the overall hazard rate of HCM, both including and excluding non-fatal endpoints, and confi-

Table 3.5: Data observation with simulation mean and confidence interval (CI,95%) of hazard rates of SCA and heart failure, and the overall hazard rate of HCM, are shown both including (inc.) and excluding (exc.) non-fatal endpoints for 20–59 years old. For the abbreviations used in the headings, please see the caption of Table 3.3. Note that the numerics are rounded to six decimal places.

Assumption	Ages	Data & Simulation	SCA		HF		Stroke	HCM	
			(exc. ReCA)	(inc. ReCA)	(exc. HT)	(inc. HT)		(exc. non-fatal)	(inc. non-fatal)
Assumption 3.4	20–29	Data Observation	0.003566	0.009509	0.001486	0.005051	0.000000	0.005349	0.014857
		Simulation Mean	0.003567	0.009506	0.001485	0.005051	0.000000	0.005351	0.014866
		(CI,95%)	(0.001783,0.005646)	(0.006240,0.012777)	(0.000297,0.002971)	(0.002674,0.007429)	(0.000000,0.000000)	(0.002971,0.008023)	(0.010994,0.019017)
	30–59	Data Observation	0.002361	0.003056	0.002361	0.005139	0.000278	0.005556	0.009028
		Simulation Mean	0.002369	0.003050	0.002363	0.005136	0.000279	0.005561	0.009021
		(CI,95%)	(0.001250,0.003611)	(0.001806,0.004306)	(0.001250,0.003611)	(0.003611,0.006806)	(0.000000,0.000694)	(0.003889,0.007361)	(0.006944,0.011250)
Assumption 3.5	20–59	Data Observation	0.002361	0.003056	0.002361	0.005139	0.000278	0.005556	0.009028
		Simulation Mean	0.002369	0.003050	0.002363	0.005136	0.000279	0.005561	0.009021
		(CI,95%)	(0.001250,0.003611)	(0.001806,0.004306)	(0.001250,0.003611)	(0.003611,0.006806)	(0.000000,0.000694)	(0.003889,0.007361)	(0.006944,0.011250)
Assumption 3.6	20–59	Data Observation	0.002210	0.003721	0.002093	0.005233	0.000233	0.005117	0.009769
		Simulation Mean	0.002204	0.003732	0.002089	0.005232	0.000230	0.005119	0.009776
		(CI,95%)	(0.001279,0.003256)	(0.002442,0.005117)	(0.001163,0.003140)	(0.003838,0.006862)	(0.000000,0.000581)	(0.003721,0.006629)	(0.007676,0.011979)

Table 3.6: Annualized hazard rates of non-fatal HCM (see Section 3.9.3).

Ages	n_x	E_x^c	Non-Fatal HCM
7–29	474	3365.4	0.009509
30–59	1000	7200	0.003472
60–91	428	2482.4	0.001611

dence intervals, under the three assumptions.

(a) When non-fatal HCM-related endpoints are excluded, which is most relevant here, the hazard rates under all three assumptions were similar, and under Assumption 3.4 they were also similar in the two age groups.

(b) When non-fatal HCM-related endpoints are included, the hazard rates were considerably higher, and under Assumption 3.4 were higher at ages 20–29 than ages 30–59. The latter difference was mainly due to SCA, with about 2/3 of occurrences being non-fatal.

(c) In total, the hazard rate of HCM-related death between ages 20 and 60 seems to lie between 0.005 and 0.0055 per year.

(d) The mean hazard rates from the bootstrapped samples in Table 3.5 were in all cases very close to the estimated hazard rates.

Based on the analysis in this section, the hazard rate of fatal HCM is assumed as 0.0055 per annum for all ages as a baseline assumption in our model. See also Table 3.6 for the hazard rates of non-fatal HCM per annum at all ages.

3.9.8 The Survival Function of the HCM-Related Hazard Rates

Consider the probability of a person age x surviving to age $x+s$ from any HCM-related endpoint(s) is denoted by ${}_sp_x$ and is given in terms of the hazard rate at age $x+s$ of any HCM-related endpoint(s) denoted by ρ_{x+s} as:

$$ {}_sp_x = \exp\left(-\int_0^s \rho_{x+w}dw\right), \tag{3.29}$$

in which ${}_sp_x + {}_sq_x = 1$. If ρ_{x+w} is constant and equal to ρ_x in the interval $w \in (0, s)$ then,

$$ {}_sp_x = \exp\left(-s\rho_x\right). \tag{3.30}$$

We use equation (3.30), with the 10,000 simulated hazard rates in $\tilde{\boldsymbol{\rho}}_{\boldsymbol{x}}$ (equation (3.27)) to obtain 10,000 simulated values of ${}_sp_x$ consistent with our assumed Poisson distribution of the number of HCM-related events. This allows us to calculate the means and 95% confidence intervals of the survival function associated with HCM-related hazard rates. The survival function from age 20, and its simulated confidence intervals (CI, 95%) are shown under Assumption 3.4 (Table 3.7 and Figure 3.6); under Assumption 3.5 (Table 3.8 and Figure 3.7); under Assumption 3.6 (Table 3.9 and Figure 3.8). In the figures, the survival probabilities are also compared with that of $q_x = 0.01$ at all integer $x \geq 0$ (the baseline assumption of the annual HCM-related mortality in Howard (2014), see Section 1.3).

3.10 Parametrising the Model IV: Mortality from All Other Causes

We use the males and females mortality rates reported in Life Tables, United States (US), 2013 (Arias et al. 2017) to quantify the hazard rates of all other causes of mortality in our model. This work gives probabilities q_x that a person age x will die before $x + 1$ (for non-negative integers x). To obtain hazard rates, we use the approximation $-\log(1 - q_x) \approx \mu_{x+0.5}$, valid for small q_x. The log-hazard rates, obtained by linear interpolation at time step 0.0005 years, are shown in Figure 3.9. Strictly, we ought to remove HCM-related mortality from the population hazard rate, since HCM is represented by a separate event in our model. We have not done so, as the effect would be small, and for our purposes the results are conservative.

3.11 Occupancy Probabilities

We first calculate the prevalence rates of HCM sub-populations at birth under the baseline assumptions (Table 3.10) as follows:

$$\frac{0.002}{{}_{20}p_0^{100} + {}_{20}p_0^{200} + {}_{20}p_0^{300} + {}_{20}p_0^{400}} \approx 0.00226, \tag{3.31}$$

respectively ${}_0p_0^{000} = p_{00} \approx (1-0.00226) = 0.99774$. Then, after age zero, we present the occupancy probabilities, obtained by the numerical solution of the Kolmogorov forward equations (see Sections 3.4 and 3.5) at time step 0.0005 years, in each state in the epidemiological model under the baseline assumptions in Figure 3.10:

- Females have higher occupancy probabilities in state 0, no events (the upper left plot) than males because they are less likely to develop late-onset of HCM (Section 3.8.2) and to die by all-cause of mortality than males (Section 3.10).

- Up to age 30, the curve of the occupancy probabilities in state 1, non-fatal HCM (the upper right plot) is upward steep because the transition intensity of non-fatal HCM between ages 7–29 is at its highest rate, see Table 3.6. The occupancy probabilities in state 2, other dead (the lower left plot) and state 3, fatal HCM (the lower right plot) are also consistent with the assumed transition intensities, see Figure 3.9 and Section 3.9.7.

We will replicate the sensitivity assumptions presented in this chapter for the adverse selection costs in Chapter 9 since our ultimate interest is to measure the insurance costs under adverse selection.

Table 3.7: The survival index of HCM for 20 years-old ($_sp_{20}$) under Assumption 3.4.

s (yr)	Data & Simulation	The Survival Probabilities of HCM (exc. non-fatal)				The Survival Probabilities of HCM (inc. non-fatal)			
		20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$	20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$
1	Data Observation	0.994666	0.942668	0.891726	0.843536	0.985253	0.854193	0.780456	0.713085
	Simulation Mean	0.994664	0.942714	0.891760	0.843624	0.985246	0.854309	0.780671	0.713469
	(CI,95%)	(0.992009,0.997033)	(0.917920,0.965211)	(0.863788,0.918852)	(0.807030,0.879127)	(0.981163,0.989066)	(0.818932,0.887958)	(0.742709,0.816720)	(0.668546,0.756457)
2	Data Observation	0.989360	0.937446	0.886786	0.838863	0.970723	0.846516	0.773442	0.706677
	Simulation Mean	0.989358	0.937487	0.886823	0.838959	0.970714	0.846639	0.773671	0.707080
	(CI,95%)	(0.984082,0.994075)	(0.912883,0.959732)	(0.857836,0.914527)	(0.801283,0.875453)	(0.962680,0.978252)	(0.811415,0.880174)	(0.734972,0.810747)	(0.661401,0.750830)
3	Data Observation	0.984082	0.932252	0.881873	0.834216	0.956407	0.838908	0.766491	0.700326
	Simulation Mean	0.984082	0.932290	0.881913	0.834321	0.956401	0.839038	0.766735	0.700750
	(CI,95%)	(0.976219,0.991125)	(0.907651,0.954678)	(0.852326,0.910694)	(0.795185,0.872176)	(0.944546,0.967555)	(0.804011,0.872580)	(0.727388,0.804418)	(0.654544,0.745531)
4	Data Observation	0.978833	0.927087	0.876987	0.829594	0.942303	0.831369	0.759603	0.694032
	Simulation Mean	0.978835	0.927122	0.877031	0.829709	0.942302	0.831506	0.759862	0.694477
	(CI,95%)	(0.968418,0.988185)	(0.902296,0.949917)	(0.846545,0.906659)	(0.789420,0.868911)	(0.926753,0.956976)	(0.796165,0.864905)	(0.720198,0.798434)	(0.647345,0.740062)
5	Data Observation	0.973612	0.921951	0.872128	0.824998	0.928407	0.823897	0.752776	0.687794
	Simulation Mean	0.973618	0.921984	0.872177	0.825122	0.928416	0.824043	0.753051	0.688261
	(CI,95%)	(0.960680,0.985253)	(0.896999,0.945076)	(0.841034,0.902460)	(0.783940,0.865630)	(0.909295,0.946512)	(0.788681,0.857633)	(0.712886,0.792048)	(0.640431,0.734675)
6	Data Observation	0.968418	0.916843	0.867297	0.820427	0.914715	0.816493	0.746011	0.681613
	Simulation Mean	0.968431	0.916875	0.867350	0.820562	0.914738	0.816648	0.746302	0.682101
	(CI,95%)	(0.953003,0.982329)	(0.891260,0.940465)	(0.835101,0.898711)	(0.778655,0.861867)	(0.892167,0.936163)	(0.781182,0.850559)	(0.705131,0.786329)	(0.633454,0.729389)
7	Data Observation	0.963252	0.911764	0.862492	0.815882	0.901226	0.809155	0.739306	0.675487
	Simulation Mean	0.963272	0.911795	0.862551	0.816028	0.901266	0.809320	0.739615	0.675997
	(CI,95%)	(0.945388,0.979415)	(0.885979,0.936271)	(0.829370,0.894749)	(0.773059,0.858400)	(0.875361,0.925927)	(0.773299,0.843479)	(0.697779,0.780405)	(0.626586,0.724140)
8	Data Observation	0.958114	0.906712	0.857713	0.811362	0.887935	0.801883	0.732662	0.669417
	Simulation Mean	0.958143	0.906743	0.857779	0.811519	0.887996	0.802059	0.732989	0.669949
	(CI,95%)	(0.937834,0.976509)	(0.880508,0.931601)	(0.823538,0.891201)	(0.767496,0.854593)	(0.858871,0.915803)	(0.765817,0.836550)	(0.690291,0.774596)	(0.619757,0.718929)
9	Data Observation	0.953003	0.901689	0.852961	0.806867	0.874841	0.794676	0.726077	0.663401
	Simulation Mean	0.953043	0.901720	0.853034	0.807036	0.874925	0.794865	0.726423	0.663956
	(CI,95%)	(0.930340,0.973612)	(0.874749,0.927286)	(0.817697,0.887125)	(0.762124,0.851140)	(0.842692,0.905790)	(0.758152,0.829981)	(0.683285,0.768488)	(0.613063,0.713755)
10	Data Observation	0.947920	0.896694	0.848236	0.802397	0.861939	0.787534	0.719552	0.657438
	Simulation Mean	0.947971	0.896726	0.848316	0.802578	0.862051	0.787735	0.719916	0.658017
	(CI,95%)	(0.922906,0.970723)	(0.869233,0.922844)	(0.812202,0.883411)	(0.756919,0.848043)	(0.826818,0.895886)	(0.750508,0.823301)	(0.675979,0.762576)	(0.606340,0.708619)

Table 3.8: The survival index of HCM for 20 years-old ($_sp_{20}$) under Assumption 3.5.

s (yr)	Data & Simulation	The Survival Probabilities of HCM (exc. non-fatal)				The Survival Probabilities of HCM (inc. non-fatal)			
		20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$	20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$
1	Data Observation	0.994460	0.940719	0.889882	0.841792	0.991013	0.905466	0.827304	0.755889
	Simulation Mean	0.994455	0.940703	0.889923	0.841947	0.991020	0.905598	0.827642	0.756491
	(CI,95%)	(0.992666,0.996119)	(0.922219,0.958124)	(0.856772,0.921579)	(0.795969,0.886428)	(0.988813,0.993080)	(0.883601,0.926456)	(0.789583,0.864302)	(0.705570,0.806317)
2	Data Observation	0.988950	0.935507	0.884952	0.837128	0.982106	0.897328	0.819869	0.749095
	Simulation Mean	0.988941	0.935494	0.885002	0.837298	0.982122	0.897478	0.820231	0.749726
	(CI,95%)	(0.985386,0.992252)	(0.915456,0.954405)	(0.850488,0.918002)	(0.790132,0.882987)	(0.977751,0.986207)	(0.873716,0.920044)	(0.780750,0.858320)	(0.697676,0.800737)
3	Data Observation	0.983471	0.930324	0.880049	0.832491	0.973280	0.889264	0.812500	0.742363
	Simulation Mean	0.983458	0.930314	0.880108	0.832674	0.973304	0.889432	0.812888	0.743023
	(CI,95%)	(0.978159,0.988401)	(0.908742,0.950701)	(0.844251,0.914439)	(0.784337,0.879560)	(0.966813,0.979382)	(0.863942,0.913677)	(0.772016,0.852381)	(0.689871,0.795196)
4	Data Observation	0.978023	0.925170	0.875173	0.827878	0.964533	0.881272	0.805198	0.735691
	Simulation Mean	0.978006	0.925164	0.875243	0.828077	0.964568	0.881459	0.805611	0.736381
	(CI,95%)	(0.970985,0.984565)	(0.902077,0.947011)	(0.838059,0.910890)	(0.778584,0.876146)	(0.955997,0.972604)	(0.854277,0.907354)	(0.763379,0.846482)	(0.682154,0.789693)
5	Data Observation	0.972604	0.920044	0.870325	0.823292	0.955865	0.873352	0.797962	0.729080
	Simulation Mean	0.972586	0.920044	0.870405	0.823506	0.955910	0.873558	0.798400	0.729799
	(CI,95%)	(0.963864,0.980743)	(0.895461,0.943335)	(0.831913,0.907354)	(0.772874,0.872746)	(0.945303,0.965874)	(0.844720,0.901075)	(0.754840,0.840624)	(0.674523,0.784228)
6	Data Observation	0.967216	0.914947	0.865503	0.818731	0.947274	0.865503	0.790790	0.722527
	Simulation Mean	0.967196	0.914952	0.865594	0.818961	0.947332	0.865730	0.791255	0.723276
	(CI,95%)	(0.956794,0.976937)	(0.888894,0.939674)	(0.825811,0.903833)	(0.767206,0.869358)	(0.934728,0.959189)	(0.835270,0.894839)	(0.746395,0.834806)	(0.666977,0.778801)
7	Data Observation	0.961858	0.909878	0.860708	0.814195	0.938761	0.857725	0.783684	0.716034
	Simulation Mean	0.961837	0.909889	0.860811	0.814442	0.938832	0.857973	0.784175	0.716813
	(CI,95%)	(0.949777,0.973145)	(0.882374,0.936027)	(0.819755,0.900325)	(0.761579,0.865984)	(0.924271,0.952551)	(0.825926,0.888647)	(0.738045,0.829029)	(0.659515,0.773411)
8	Data Observation	0.956529	0.904837	0.855940	0.809684	0.930324	0.850016	0.776640	0.709599
	Simulation Mean	0.956508	0.904855	0.856055	0.809948	0.930409	0.850286	0.777159	0.710409
	(CI,95%)	(0.942812,0.969368)	(0.875903,0.932394)	(0.813743,0.896830)	(0.755994,0.862623)	(0.913931,0.945959)	(0.816686,0.882497)	(0.729789,0.823292)	(0.652137,0.768059)
9	Data Observation	0.951229	0.899824	0.851197	0.805198	0.921963	0.842377	0.769661	0.703222
	Simulation Mean	0.951210	0.899849	0.851326	0.805479	0.922063	0.842669	0.770206	0.704062
	(CI,95%)	(0.935897,0.965605)	(0.869479,0.928775)	(0.807775,0.893349)	(0.750449,0.859275)	(0.903707,0.939413)	(0.807550,0.876390)	(0.721625,0.817594)	(0.644842,0.762744)
10	Data Observation	0.945959	0.894839	0.846482	0.800737	0.913677	0.834806	0.762744	0.696902
	Simulation Mean	0.945941	0.894872	0.846623	0.801036	0.913793	0.835121	0.763317	0.697773
	(CI,95%)	(0.929033,0.961858)	(0.863102,0.925170)	(0.801850,0.889882)	(0.744945,0.855940)	(0.893597,0.932912)	(0.798516,0.870325)	(0.713552,0.811936)	(0.637628,0.757465)

Table 3.9: The survival index of HCM for 20 years-old ($_sp_{20}$) under Assumption 3.6.

s (yr)	Data & Simulation	The Survival Probabilities of HCM (exc. non-fatal)				The Survival Probabilities of HCM (inc. non-fatal)			
		20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$	20 years-old $_sp_{20}$	at 30 years-old $_{s+10}p_{20}$	at 40 years-old $_{s+20}p_{20}$	at 50 years-old $_{s+30}p_{20}$
1	Data Observation	0.994896	0.945267	0.898114	0.853313	0.990279	0.898114	0.814527	0.738720
	Simulation Mean	0.994894	0.945279	0.898190	0.853496	0.990272	0.898109	0.814618	0.738974
	(CI,95%)	(0.993393,0.996285)	(0.929677,0.959890)	(0.870048,0.924824)	(0.814243,0.891040)	(0.988093,0.992354)	(0.876548,0.919035)	(0.777595,0.851133)	(0.689813,0.788248)
2	Data Observation	0.989818	0.940443	0.893530	0.848958	0.980652	0.889383	0.806609	0.731539
	Simulation Mean	0.989815	0.940458	0.893615	0.849154	0.980641	0.889384	0.806713	0.731811
	(CI,95%)	(0.986830,0.992585)	(0.923535,0.956325)	(0.864299,0.921389)	(0.808864,0.887730)	(0.976328,0.984766)	(0.866111,0.912008)	(0.768336,0.844625)	(0.681599,0.782221)
3	Data Observation	0.984766	0.935643	0.888970	0.844625	0.971119	0.880737	0.798768	0.724427
	Simulation Mean	0.984762	0.935663	0.889064	0.844834	0.971104	0.880745	0.798886	0.724719
	(CI,95%)	(0.980310,0.988898)	(0.917433,0.952772)	(0.858589,0.917967)	(0.803519,0.884432)	(0.964702,0.977236)	(0.855798,0.905034)	(0.759187,0.838167)	(0.673483,0.776240)
4	Data Observation	0.979740	0.930867	0.884432	0.840314	0.961678	0.872175	0.791003	0.717385
	Simulation Mean	0.979736	0.930893	0.884536	0.840537	0.961661	0.872191	0.791136	0.717697
	(CI,95%)	(0.973833,0.985224)	(0.911371,0.949233)	(0.852916,0.914557)	(0.798211,0.881147)	(0.953216,0.969764)	(0.845608,0.898114)	(0.750148,0.831758)	(0.665464,0.770304)
5	Data Observation	0.974739	0.926116	0.879918	0.836025	0.952329	0.863697	0.783313	0.710411
	Simulation Mean	0.974736	0.926147	0.880032	0.836262	0.952310	0.863720	0.783462	0.710743
	(CI,95%)	(0.967399,0.981565)	(0.905350,0.945707)	(0.847281,0.911159)	(0.792937,0.877874)	(0.941866,0.962349)	(0.835539,0.891247)	(0.741216,0.825398)	(0.657540,0.764414)
6	Data Observation	0.969764	0.921389	0.875427	0.831758	0.943071	0.855300	0.775698	0.703504
	Simulation Mean	0.969762	0.921427	0.875552	0.832009	0.943052	0.855333	0.775863	0.703858
	(CI,95%)	(0.961007,0.977918)	(0.899368,0.942194)	(0.841683,0.907775)	(0.787698,0.874613)	(0.930651,0.954991)	(0.825590,0.884432)	(0.732390,0.819087)	(0.649711,0.758570)
7	Data Observation	0.964815	0.916686	0.870959	0.827513	0.933903	0.846986	0.768157	0.696665
	Simulation Mean	0.964814	0.916731	0.871094	0.827778	0.933885	0.847029	0.768339	0.697040
	(CI,95%)	(0.954658,0.974286)	(0.893426,0.938694)	(0.836122,0.904403)	(0.782494,0.871364)	(0.919569,0.947689)	(0.815760,0.877670)	(0.723669,0.812824)	(0.641975,0.752769)
8	Data Observation	0.959890	0.912008	0.866514	0.823289	0.924824	0.838752	0.760690	0.689893
	Simulation Mean	0.959892	0.912059	0.866660	0.823570	0.924808	0.838806	0.760889	0.690289
	(CI,95%)	(0.948350,0.970667)	(0.887523,0.935207)	(0.830598,0.901043)	(0.777324,0.868127)	(0.908620,0.940443)	(0.806046,0.870959)	(0.715052,0.806609)	(0.634331,0.747014)
9	Data Observation	0.954991	0.907353	0.862091	0.819087	0.915834	0.830598	0.753295	0.683186
	Simulation Mean	0.954995	0.907412	0.862250	0.819383	0.915820	0.830663	0.753512	0.683604
	(CI,95%)	(0.942085,0.967061)	(0.881660,0.931734)	(0.825110,0.897696)	(0.772188,0.864903)	(0.897801,0.933252)	(0.796449,0.864299)	(0.706538,0.800442)	(0.626778,0.741302)
10	Data Observation	0.950117	0.902722	0.857691	0.814906	0.906931	0.822523	0.745972	0.676545
	Simulation Mean	0.950124	0.902789	0.857862	0.815218	0.906921	0.822601	0.746207	0.676985
	(CI,95%)	(0.935860,0.963469)	(0.875834,0.928273)	(0.819659,0.894362)	(0.767086,0.861690)	(0.887111,0.926116)	(0.786965,0.857691)	(0.698125,0.794321)	(0.619314,0.735634)

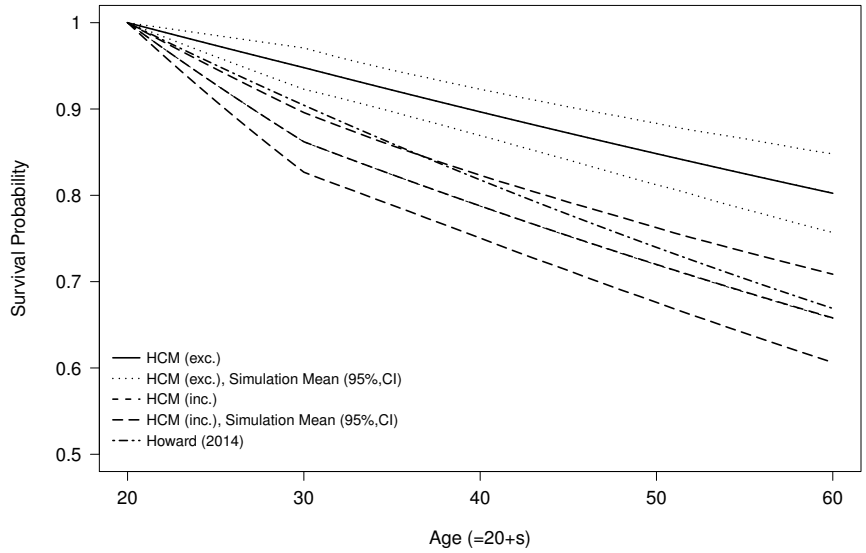


Figure 3.6: The survival function from age 20 ($_sp_{20}$) associated with HCM-related hazard rates (including and excluding non-fatal HCM) under Assumption 3.4.

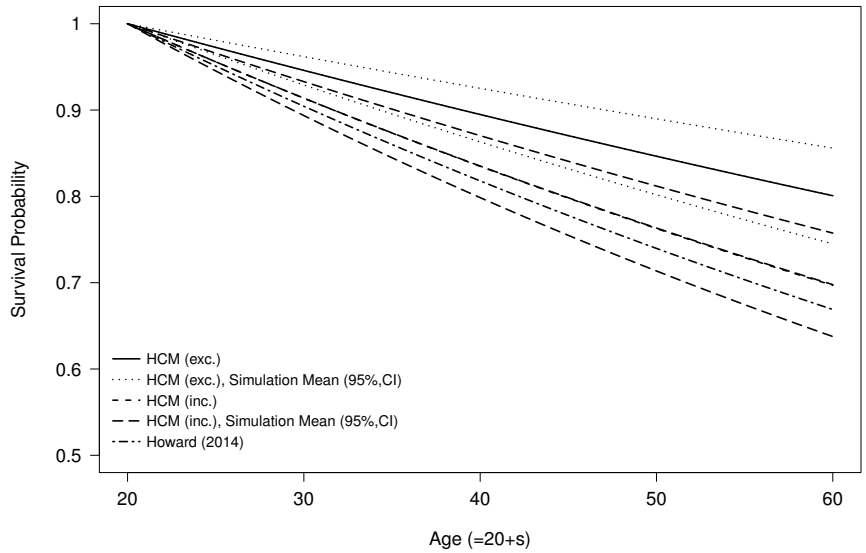


Figure 3.7: The survival function from age 20 ($_sp_{20}$) associated with HCM-related hazard rates (including and excluding non-fatal HCM) under Assumption 3.5.

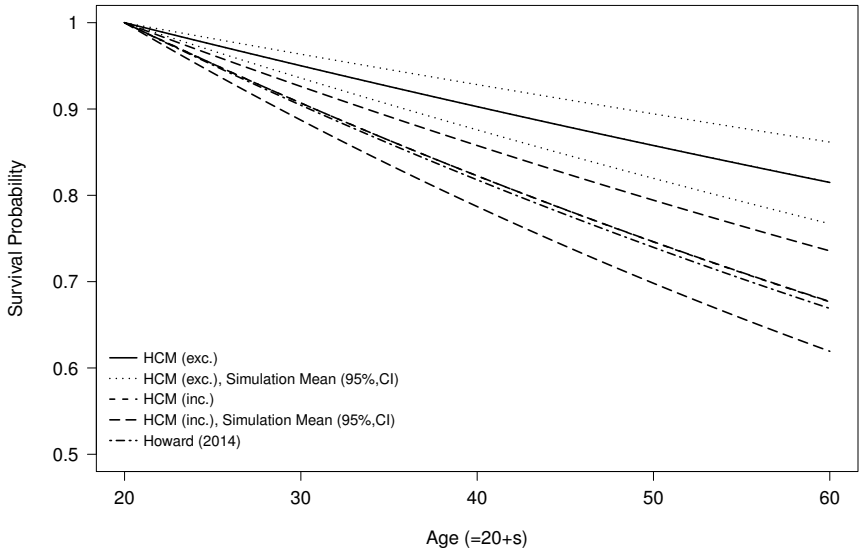


Figure 3.8: The survival function from age 20 ($_sp_{20}$) associated with HCM-related hazard rates (including and excluding non-fatal HCM) under Assumption 3.6.

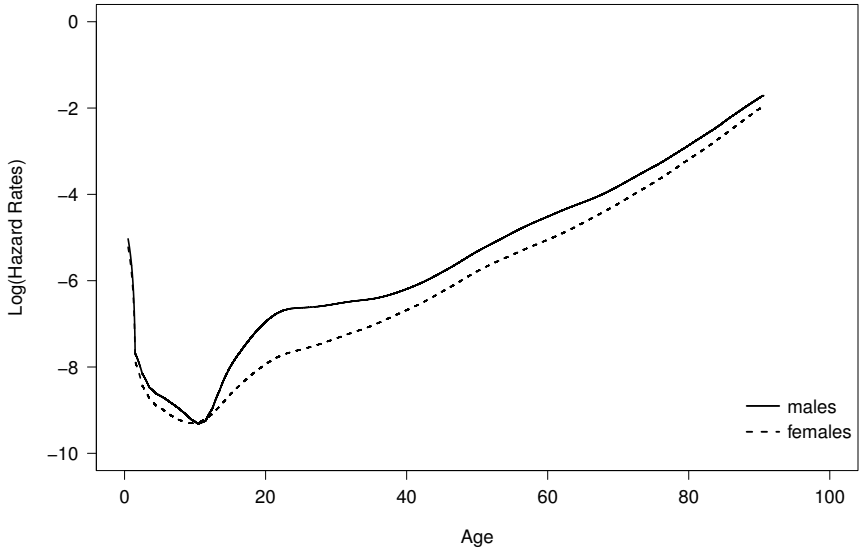


Figure 3.9: Estimated hazard rates of all-cause mortality from the reported mortality rates in Life Tables, United States (US), 2013 (Arias et al. 2017).

Table 3.10: Baseline assumptions for the epidemiological model parameters.

Prevalence of non-HCM mutations in the general population at age 20	0.998	Section 3.7.1
Prevalence of HCM mutations in the general population at age 20	0.002	Section 3.7.1
Prevalence of known early-onset mutations in the HCM population at birth	0.5	Section 3.7.2
Prevalence of known late-onset mutations in the HCM population at birth	0.1667	Section 3.7.2
Prevalence of unknown early-onset mutations in the HCM population at birth	0.25	Section 3.7.2
Prevalence of unknown late-onset mutations in the HCM population at birth	0.0833	Section 3.7.2
Penetrance of early-onset HCM at age 20	100%	Section 3.8.1
Penetrance of late-onset HCM at ages 20–70	Figure 3.4	Section 3.8.2
Hazard rate of fatal HCM per annum for all ages	0.0055	Section 3.9.7
Hazard rate of non-fatal HCM per annum for all ages	Table 3.6	Section 3.9.7
Hazard rate of all other death per annum for all ages	Figure 3.9	Section 3.10

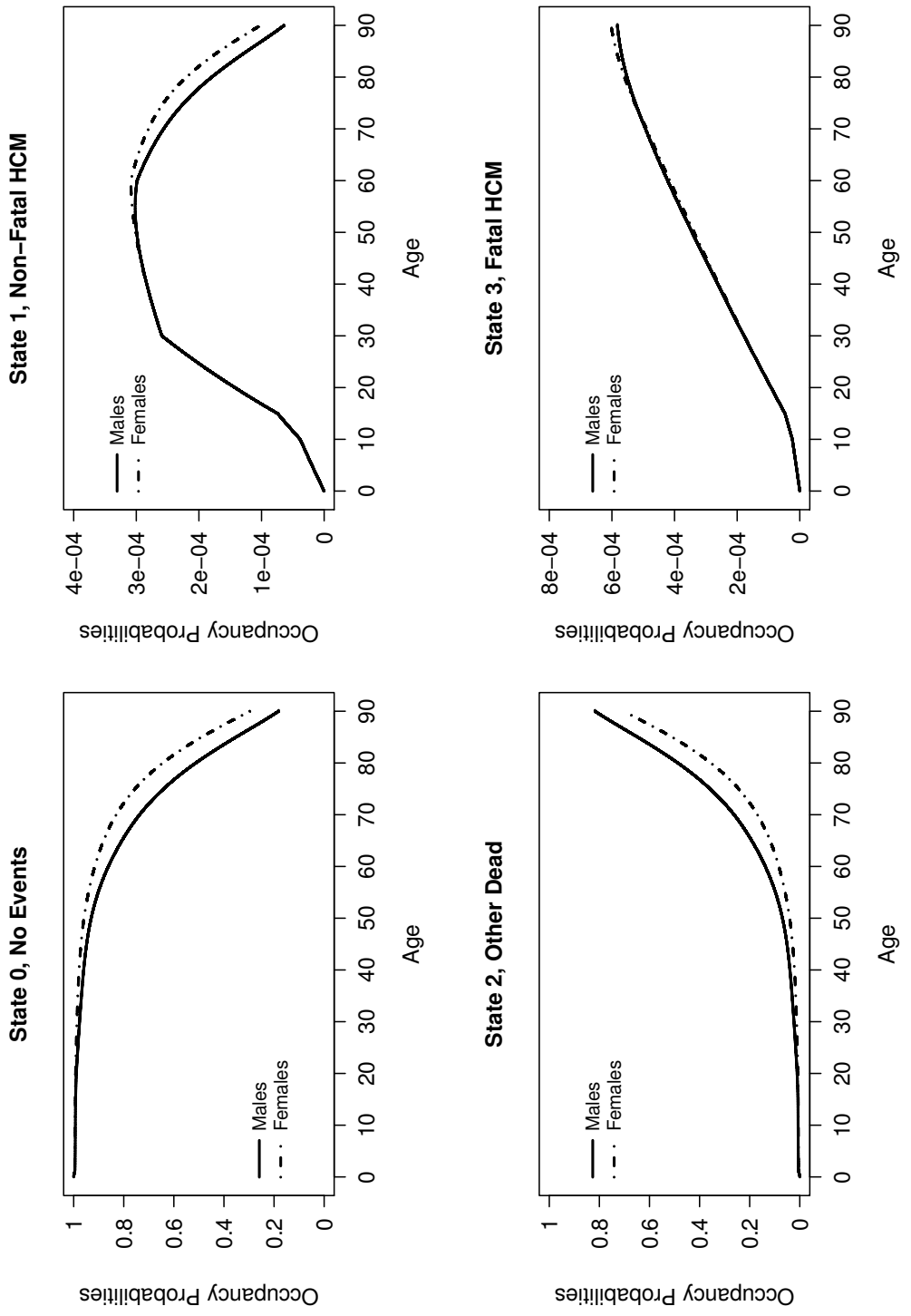


Figure 3.10: Occupancy probabilities in each state in the epidemiological model under baseline assumptions for males and females separately.

Chapter 4

Genetic Testing in Hypertrophic Cardiomyopathy (HCM)

4.1 Introduction

We described the genetic substrate of HCM in Chapter 2, and left the methodology of genetic testing in HCM for this chapter. In Chapter 1, we stated that this study *ultimately* attempts to model the impact of genetic testing in HCM for the life insurance market. We do so, starting in Chapter 5, by specifying 'the testing model' of HCM adding genetic testing states to the epidemiological model in Chapter 3, and then in Chapter 7, by specifying 'the adverse selection model' of HCM for a life insurance market by adding insurance states to the testing model. The mathematical foundations of the testing and adverse selection models are established not only by adding more states to the epidemiological model, but also by specifying models of testing and insurance purchase behaviour on the part of individuals in HCM families. (Note that an HCM family is a new term in this study describing a mutation carrier family member and relatives of this member of the family).

The epidemiological model, in Chapter 3, modelled the life history of a single individual, mainly exposed to the risk of adverse HCM-related events, such as non-fatal and fatal HCM. In this chapter, we explain the nature of genetic testing in HCM families. In Section 4.2, we introduce the form of genetic testing, called *cascade genetic testing*, which is widely used in HCM. In Section 4.3, we present the

uptake rates of cascade genetic testing in HCM, reported in medical studies and we show how we incorporate these rates into our study. Section 4.4 discusses how we can represent the nature of cascade genetic testing in HCM through a mathematical model.

4.2 Cascade Genetic Testing

The form of genetic testing most associated with HCM is *cascade genetic testing*. This begins when a person is clinically diagnosed with HCM. If they are a member of a family with no previously known cases of HCM, this event reveals the presence of a mutation in the family. Cascade genetic testing means that genetic testing is offered to the affected person, their parents, their siblings and their children (known as 'first-degree relatives'). Depending on the results, genetic testing may then be offered to more distant relatives, for example the siblings of an affected parent or the children of an affected sibling.

Hence the process of genetic testing can 'cascade' through an extended family tree. Elliott et al. (2014) describe the process.

(a) The person in a family first clinically diagnosed with HCM is called the *index patient* or *proband*. In our model, for simplicity, we assume that the first family member to manifest HCM always initiates cascade genetic testing and is the proband. They are said to be phenotype-positive (phenotype:+), (the *phenotype* is the physical manifestation of a gene variety). The proband is advised to undergo genetic counselling, then genetic testing. They are not obliged to undergo either.

(b) The proband may be found to carry a known HCM-related mutation, denoted by (phenotype:+, genotype:+). Then, all first-degree relatives may be advised to undergo counselling and genetic testing.

    (i) Relatives who are found to be genetically and clinically affected by HCM (genotype:+, phenotype:+) or genetically, but not clinically, affected by

HCM (genotype:+, phenotype:−) are followed up clinically. Genetic testing may be offered to their first-degree relatives in cascade fashion as described above.

(ii) Relatives who are found to be genotype-negative are evaluated as not at risk of HCM.

(c) The proband may be found not to carry a known HCM-related mutation, denoted by (phenotype:+, genotype:−). It means that the genetic substrate of these patients is 'unknown'; either genetic testing cannot detect a known HCM-related mutation or finds a variant of unknown significance (VUS). Due to the presumption in Section 2.3.1.2, we assume that they carry a mutation yet to be identified. Genetic testing cannot be carried out on their relatives, who are still at risk of HCM, however. Therefore, cascade clinical screening may be offered to the relatives. Those found to be phenotype-negative are advised to repeat clinical screening at intervals, as they do not have the clear indication of a negative genetic test for a known mutation. Those found to be phenotype-positive can receive appropriate treatment.

(d) If the gene mutation in the proband is found to be a variant of unknown/uncertain significance (VUS), segregation analysis is recommended if possible, and then either cascade genetic testing or cascade clinical screening is advised for the relatives.

(e) Points (b) and (c) above clarify that a proband can be detected by:

(i) Variants which are known and known to be associated with a phenotype. In this study, they are referred as 'known mutations'. Any person carrying such a mutation is referred to 'a carrier of a known HCM mutation'.

(ii) Variants which are presumed to exist because of observed phenotype but which have not been detected. Or, variants which are known but whose association with a phenotype is uncertain, called variant of unknown significance (VUS). In this study, they are all referred to as 'unknown

mutations'. A proband tested with an unknown HCM mutation is referred to 'a carrier of an unknown HCM mutation'. The relatives of such a proband, as pointed out above, do not undergo genetic testing, but they are clinically followed up at intervals.

4.3 Uptake Rates of Genetic Testing

Not all relatives offered genetic testing agree to undergo it. We call the proportion who do agree to undertake it the 'uptake rate' of genetic testing. In what follows we discuss significant aspects of the uptake rates of cascade genetic testing.

(a) First and foremost, we need a proband to start cascade genetic testing in the family.

(b) Secondly, we define the offer rate to be the proportion of new probands, diagnosed as phenotype +, who are referred to a genetics clinic to initiate cascade testing of their relatives. Khouzam et al. (2015) conducted a questionnaire study aiming to determine the factors influencing the uptake of genetic services in HCM in the USA. The study was limited to first-degree relatives and did not identify probands and their relatives separately. It included 306 persons; of whom, 264 replied to the question of whether they were either discussed or offered to take-up genetic testing. 210 of 264 (80%) claimed that they were offered genetic testing. The offer rate seems to be less than 100%. More results from this questionnaire study will be shown in Table 4.1.

(c) Thirdly, we define the take-up rate in probands to be the proportion of the probands who were offered genetic testing that did eventually take up the testing. There are no strong data available to quantify this rate. Geelen et al. (2012) suggest that the take-up rate in probands is less than 100%. However, this study does not estimate this take-up rate. The study sample is also very small (four HCM families) and its aim was to find out the behavioural reasons for not taking up genetic testing in either probands or their relatives.

(d) Based on these observations, all probands are assumed to be offered and to take up the genetic testing in our model as a conservative assumption.

Actually, the main purpose of cascade genetic testing in HCM is to detect genetically affected, but clinically undiagnosed, or asymptomatic, at-risk relatives of the proband to improve their treatment by early care. Then the term 'the uptake rate of testing' will be mostly used here to refer to the proportion of the at-risk relatives of a proband who take up the offer of testing.

4.3.1 Uptake Rates of Genetic Testing in Medical Studies

We show our literature review on the uptake rates of the testing in or related to HCM in Table 4.1. However, the studies generally reported uptake rates without a follow-up period between the detection of the mutation in probands and the testing of the at-risk relatives of the probands. Therefore, we checked the uptake rates of testing of at-risk relatives of a proband in genetic disorders other than HCM, with their follow-up periods, shown in Table 4.2.

4.3.2 Uptake Rates of Genetic Testing in This Study

A 'rate' in these studies (see Tables 4.1 and 4.2) refers to the cumulative probability that a person is tested between ages x and $x + s$. Based on the reported uptake rates or probabilities of being tested in medical studies, we pick a baseline constant proportion which is 50% of the at-risk relatives of the proband taking up genetic testing, during the first year since (if) a proband exists in the family (see Tables 4.1 and 4.2). There is no age restriction on taking up either genetic testing or clinical screening for the at-risk relatives of probands. If there is a proband in a family with young children, the ESC guidelines (Elliott et al. 2014) recommend to start genetic testing or clinical screening of the at-risk relatives after age 10. The guidelines also state that the first-degree relatives at ages less than 10 could take up both genetic testing and clinical screening if they have an adverse family history or they have the symptoms of HCM at these early ages. The highest ages at genetic testing were 66 years in Charron et al. (2002) and 70 years in Brook et al. (2004). The other studies

Table 4.1: Uptake rates of genetic testing in/related to HCM at-risk relatives of mutation carrier probands. FDR: First-degree relative, SDR: Second-degree relative, Follow-Up Period: Observation time between the detection of family mutation in probands and the testing of the relatives of the probands.

Population	Age (yr)	Disorder (Causing Mutant Genes)	Family Pedigree	Follow-Up Period	Uptake Rate of Genetic Testing	Reference
Up to second-degree at-risk relatives in France.	16–66	HCM (Majority with MYBPC3 or MYH7)	-	-	-	Charron et al. (2002)*
First and second-degree relatives *that included on the condition of death of the first-degree relative* in the Netherlands.	>10	HCM (Majority with MYBPC3)	Overall	1 year	38.61%	Christiaans et al. (2011)†
A sample of 306 persons. Of whom, 270 clinical HCM patients and 36 at-risk first-degree relatives in the USA.	>18	HCM	-	-	53%	Khouzam et al. (2015)‡
First and second-degree at-risk relatives in the USA.	-	HCM and DCM (A large variety)	FDRs SDRs Overall	-	51% 16% 39%	Miller et al. (2013)¶
At-risk children in 58 families in Canada. 29 families had their children tested.	<18	LQTS, HCM, and ARVC	Only children	-	66%	Christian et al. (2018)§

* In **Charron et al. (2002)**, 60 subjects had a family history with HCM. Of whom, (a) 29 were adult relatives taking up genetic counselling at ages 18–66 years old. 26 of 29 (90%) were first and 3 of 29 (10%) were second-degree relatives. 19 of 29 took up the testing. (b) 9 of 60 were couples taking up genetic counselling for their children at ages 15 months to 16 years. 4 of 9 had their two children counselled while the rest of 9 had their one child counselled (total number of children of each couple were not submitted). Therefore, this sample in total contained 13 children. However, only one couple had their child tested. (c) 22 of 60 were couples taking up prenatal counselling since these couples had family history with HCM. None of them were offered to take-up the testing.

† In **Christiaans et al. (2008)**, at-risk relatives of 97 HCM mutation carrier probands were included. Of whom, (a) 68 probands (70%) took up the counselling in the first year. Note that 99% of the relatives of these sixty-eight probands taking up the counselling took up the testing in the first year. Also, the uptake rate of the counselling was 40.4% in first and 27.5% in second-degree relatives. (b) None of the relatives of the remaining 29 (%30) took up the counselling in the first year. The relatives of 16 of 29 probands took up the counselling after the first year. The relatives of 13 of 29 probands did not take-up the counselling at all during the mean follow-up period 53 months of a range of 16–103 months. (c) The uptake of genetic counselling in HCM was not significantly related to age, gender, or family history with SCD.

‡ In **Khouzam et al. (2015)**, the results of a questionnaire study were reported. (a) The study mainly focused to determine the factors affecting uptake of genetic testing in HCM. (b) The study did not question on follow up period of the process of uptake of genetic testing in HCM. (c) The study did not find significant relationship between uptake of genetic testing and gender, nor family history with SCD.

¶ In **Miller et al. (2013)**, at-risk relatives of 40 HCM- or DCM-related mutation carrier probands were examined. Of whom, 34 (85%) were HCM- and 6 (15%) were DCM-related mutation carriers. Uptake of genetic testing was not significantly associated with either age or family history with SCD.

§ In **Christian et al. (2018)**, the high uptake rate of genetic testing compared to previous studies was explained in the study as follows: (a) mutation carrier parents already taken up the testing were preferred and (b) sample consisted of mutation carrier parents of LQTS, HCM, ARVC disorders.

Table 4.2: Uptake rates of genetic testing in other inherited disorders at-risk relatives of mutation carrier probands. FDR: First-degree relative, SDR: Second-degree relative, DR: Distant relative, P-T risk: Pre-test risk of inheriting of a known family mutation, Follow-Up Period: Observation time between the detection of family mutation in probands and the testing of the relatives of the probands.

Population	Age (yr)	Disorder (Causing Mutant Genes)	Family Pedigree	Follow-Up Period	Uptake Rate of Genetic Testing	Reference
50% and 25% P-T risk relatives in the Netherlands.	≥20	Breast/Ovarian Cancer (BRCA1/2)	50% P-T	9 months,1 year,2 year	51%,54%,58% in women	Meijers-Heijboer et al. (2000)
			50% P-T	9 months,1 year,2 year	19%,19%,24% in men	
50%, 25%, 12.5% P-T risk relatives *"in each vertical bloodline in the family pedigree"* in Manchester (MCR) and London (LDN) in the UK.	18–70	Breast/Ovarian Cancer (BRCA1/2)	FDRs, MCR	-	52.8% women; 14.7% men	Brook et al. (2004)
			FDRs, LDN	-	23.5% women; 7.4% men	
			SDRs, MCR	-	46.7% women; 16% men	
			SDRs, LDN	-	25% women; 0% men	
			DRs, MCR	-	61% women; 4.9% men	
			DRs, LDN	-	45% women; 66% men	
			Overall, MCR	mean 1.3 years (0.1–5.8)	33% (53% women;12.3% men)	
			Overall, LDN	mean 1 year (0.1–4.1)	21% (28.6% women;11.4% men)	
First and second-degree, and more distant at-risk relatives in Scotland in the UK.	<18 and ≥18	Breast/Ovarian Cancer (BRCA1/2)	FDRs	median 7 months (1–56)	39% (58% women;8% men)	Holloway et al. (2008)
			SDR&DRs	median 14 months (2–61)	26% (39% women;9% men)	
			Overall	median 37 months	32% (47% women;8% men)	
First and second-degree at-risk relatives in Spain.	≥18	Breast/Ovarian Cancer (BRCA1/2)	FDRs	median 2 months (0–113)	60% (76% women;39% men)	Sanz et al. (2010)
			SDRs	median 6 months (0–73)	28% (37% women;17% men)	
			Overall	median 3.3 months (0–113)	44% (58% women;29% men)	
100% *(those who had HNPCC tumor)*, 50%, 25% P-T risk relatives in the Netherlands.	≥18	HNPCC (MSH2, MLH1, or MSH6)	100% P-T	-	87%	Wagner et al. (2002)
			50% P-T	-	57%**	
			25% P-T	-	21%	
			Overall	mean 42 months (12–74)	50%	
50% and 25% P-T risk relatives in Australia.	≥16	Breast/Ovarian Cancer (BRCA1/2), HNPCC (MLH1, MSH2, MSH6), Cowden Syndrome (PTEN, only 1 relative).	-	2 years	40%	Suthers et al. (2006)
A systematic review study for at-risk relatives	-	Lynch Syndrome‖	FDRs	-	34–52%	Sharaf et al. (2013)
A systematic review study for at-risk relatives	-	Breast Cancer	-	-	mean 59% (25–96%)	Ropka et al. (2006)

** 41%, 58%, and 65% of 57% took up the testing in 1 year, 2, and 3 years.

‖ Lynch Syndrome is also known as HNPCC.

did not report the upper age limit of taking up genetic testing. Therefore, this study conservatively assumes the uptake ages of genetic testing of at-risk relatives in HCM to be in the range 0–70 years. Note that we have little interest after age 60 since we assume individuals leave the life insurance market at age 60.

Then, the hazard rate of the uptake of the testing at-risk relatives in our study, see Section 3.9.8, is taken into account as follows:

$$\mu_{x,z} = \begin{cases} -\log(0.5) = 0.6931472, & 0 < x < 70,\ 0 < z \leq 1. \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

where $\mu_{x,z}$ is the hazard rate for an individual who transfers from an 'untested' state to a 'tested' state. Label x is the age of the individual and z is the duration in an 'untested' state since the appearance of the proband.

In addition, the assumption above allows individuals in HCM families to take up genetic testing immediately after a proband appears in the family. We extend this assumption, which will be conservative for our purposes, by allowing children to take up genetic testing at birth if either parent is a proband. See Section 5.9.3.

Moreover, all the uptake rates described above were based on the relatives of probands who were detected with a 'known mutation' leading to the genetic disorders. In our study, if probands are detected with an 'unknown mutation', we assume that the uptake rate of testing of the at-risk relatives of these probands is always zero, because testing is not offered.

4.4 From Genetic Testing in HCM to a Mathematical Model of the Uptake of Genetic Testing in HCM

Here we consider how to reflect the nature of cascade genetic testing in a mathematical model. In Chapter 3, the epidemiological model of HCM was a multiple-state Markov model for the life history of a single individual in which transition inten-

sities depend only on the current state and age, so we could compute occupancy probabilities using the Kolmogorov forward equations. We can easily extend the state space of the epidemiological model by adding testing states. By this we mean, adding transitions into new model states, with the transition representing the event of taking up a genetic test.

In cascade genetic testing, the appearance of a proband in a family will change the testing behaviour of other family members. A mathematical model of cascade genetic testing should capture who is the proband and when the proband appeared in the family. A proband appearing in a family is an event that happens at a random time, which then modifies the behaviour of other family members. From this, we conclude the following.

(a) We must model the whole family, not just model its members separately. We have to capture the changed behaviour of family members when a proband appears, at a random time.

(b) If we model the life history of a family member as a multiple-state model (which we do) then the model is no longer Markov, for two reasons.

  (i) Transition intensities (representing uptake of genetic testing) now depend on the states occupied by other family members.

  (ii) The same transition intensities depend on duration since a proband appeared (equation (4.1)).

Without a vast increase in complication, we cannot specify even a semi-Markov model (allowing for point (b) (ii) above) representing the whole family. Our approach, described in Chapter 5, is to abandon the Kolmogorov equations in these circumstances and simulate the linked life histories of all the members of a family in which an HCM-related mutation is present.

Chapter 5

A Simulation Model of the Uptake of Genetic Testing in Hypertrophic Cardiomyopathy (HCM)

5.1 HCM and Non-HCM Families

The epidemiological model, developed in Chapter 3, was a model of the life history of a single individual. It is easy to extend this to allow for genetic testing by adding one or more transitions from untested to tested states, which we do in Section 5.2.

However, when genetic testing occurs in cascade fashion (see Section 4.2), the transition intensity from an untested state to a tested state depends on whether or not there is a proband (index case) in the family; and, duration in the untested state since (if) a proband exists in the family, see Section 4.3.2. It means that transition intensities depend on information other than time or age and the state currently occupied, so the model is not Markov. This has two major consequences, which we describe in this chapter.

1. We cannot use the Kolmogorov forward equations to obtain occupancy probabilities.

2. The additional information, upon which the transition intensities relating to genetic testing depend, relates to the life histories of other family members, namely the presence or absence of a proband in the family. We cannot now model just the life history of a single individual, we have to model, simultaneously in calendar time, the joint life histories of all the members of a family. (Note that the transition intensities of genetic testing also depend on duration in the 'untested' state after (if) a proband appears in the family.)

The approach we adopt is to simulate the life histories of all the members of a family in which one parent has an HCM-related mutation. For this purpose we define a 'family' to be a nuclear family consisting of two parents and a number of children (possibly zero). (See Section 4.1 for the definition of an HCM family in which this study models a nuclear family, or a family up to two generations). Since such families are a small minority in the general population, we adopt a two-pronged approach.

(a) For these families, which we call 'HCM families' we simulate life histories explicitly. This is time-consuming, but unavoidable. See Sections 5.4, 5.6, and 5.9. Note that these families are a mixture of individuals in HCM and non-HCM populations defined in Section 5.2. Figures 5.2 and 5.3 clarify what an HCM population and HCM family refers to in our model.

(b) For other families, the great majority, which we call 'non-HCM families' cascade genetic testing for HCM can never be initiated and life histories are Markov. Therefore, we may still model life histories in these families by the more efficient method of solving numerically the Kolmogorov equations. See Sections 5.5, 5.7, and 5.8. Note that these families are always part of non-HCM population, see also Figures 5.2 and 5.3.

We want to simulate the testing behaviour in HCM families. Before doing so, we introduce the testing model in Section 5.2. We show how we represent families in the general population and model their life histories in Section 5.3. In Sections 5.4 and 5.5, we form HCM and non-HCM families by pairing of spouses. In Sections 5.6

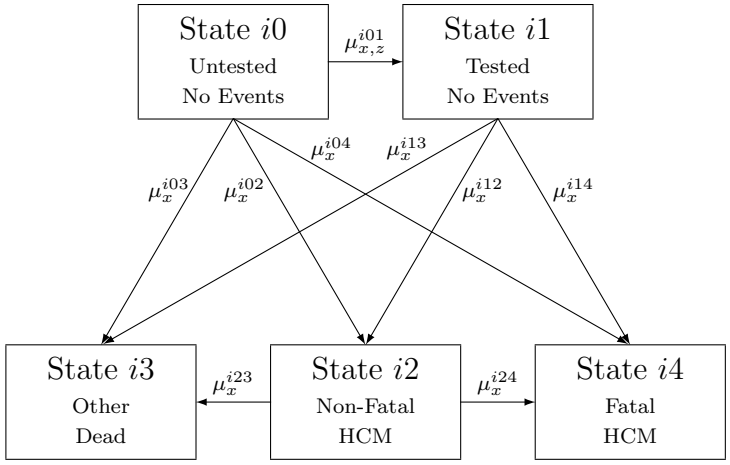


Figure 5.1: A mathematical model of uptake of genetic testing in HCM, representing the uptake of genetic testing of a person in the ith of several sub-populations defined by HCM genotype. In $\mu_{x,z}^{i01}$, z refers to duration in state $i0$ since (if) a proband exists in the family.

and 5.7, we add children to the HCM and non-HCM families so formed. In Section 5.8, we present the life histories in respect of non-HCM families. In Section 5.9, we describe the simulation algorithm to simulate testing behaviour in HCM families and show the simulation results of the life histories in respect of HCM families. We discuss the testing model in Section 5.10.

5.2 The Testing Model

Our mathematical model of genetic testing in HCM, which is similar to the epidemiological model in Figure 3.1, but including genetic testing states, is the multiple-state model shown in Figure 5.1. Therefore, the mathematical basis of the testing model is not significantly different from that of the epidemiological model introduced in Section 3.3; we have just added one more model state in all sub-populations, representing genetic testing.

The testing model contains 45 states (nine sub-populations where each sub-population contains five sub-states). The sub-populations in the epidemiological model only show that individuals either carry or do not carry an HCM-related mutation (see Section 3.2). Non-carriers of HCM-related mutations were there represented in one sub-population in non-HCM population while carriers of HCM-related

mutations were there represented in four sub-populations, based on the type of the mutation, in HCM population (see Figure 3.2).

However, in the testing model, we also divide non-carriers of HCM-related mutations into different sub-populations in non-HCM population because testing behaviour may change in all family members, both non-carriers (in non-HCM population) and carriers (in HCM population), in HCM families. For example, a non-carrier member of a family (in non-HCM population, but in HCM family) affected with a known HCM mutation will learn they are a non-carrier only after being tested. Figures 5.2 and 5.3 show all the testing model states with the nine sub-populations (note that both figures refer to the same model) in which each sub-population is classified by belonging to two kinds which are population (HCM or non-HCM population, see Figure 5.2) and family (HCM or non-HCM families, see Figure 5.3).

The new sub-populations are labelled $i = 1, 3, 5, 7$. The corresponding carrier sub-populations are now labelled $i = 2, 4, 6, 8$. Sub-populations $i = 1, 3, 5, 7$ will only contain two types of individuals:

(a) Spouses of individuals in a carrier sub-population.

(b) Non-carrier children born to a couple, one of whom is a carrier.

We need to distinguish such individuals from other non-carriers in sub-population $i = 0$ (in non-HCM population and non-HCM families) because, while their biological risks are identical, their genetic testing and (in Chapter 7) insurance-purchasing behaviour may not be. We may refer to sub-population $i = 1, 3, 5, 7$ (in non-HCM population and HCM families) as being the complements of sub-populations $i = 2, 4, 6, 8$ (in HCM population and HCM families) respectively.

In Figure 5.1, the genotype is known and fixed. In Figures 5.2 and 5.3 (we noted above that both figures refer to the same model), the genotype is unknown, which allows us to model probabilistically the life history of an individual. When we talk of modelling a family, we mean a collection of models of the kind shown in Figures 5.1, 5.2, and 5.3, one for each member of the family. The individual models in the collection are not independent because events in one model may change transition intensities in the others.

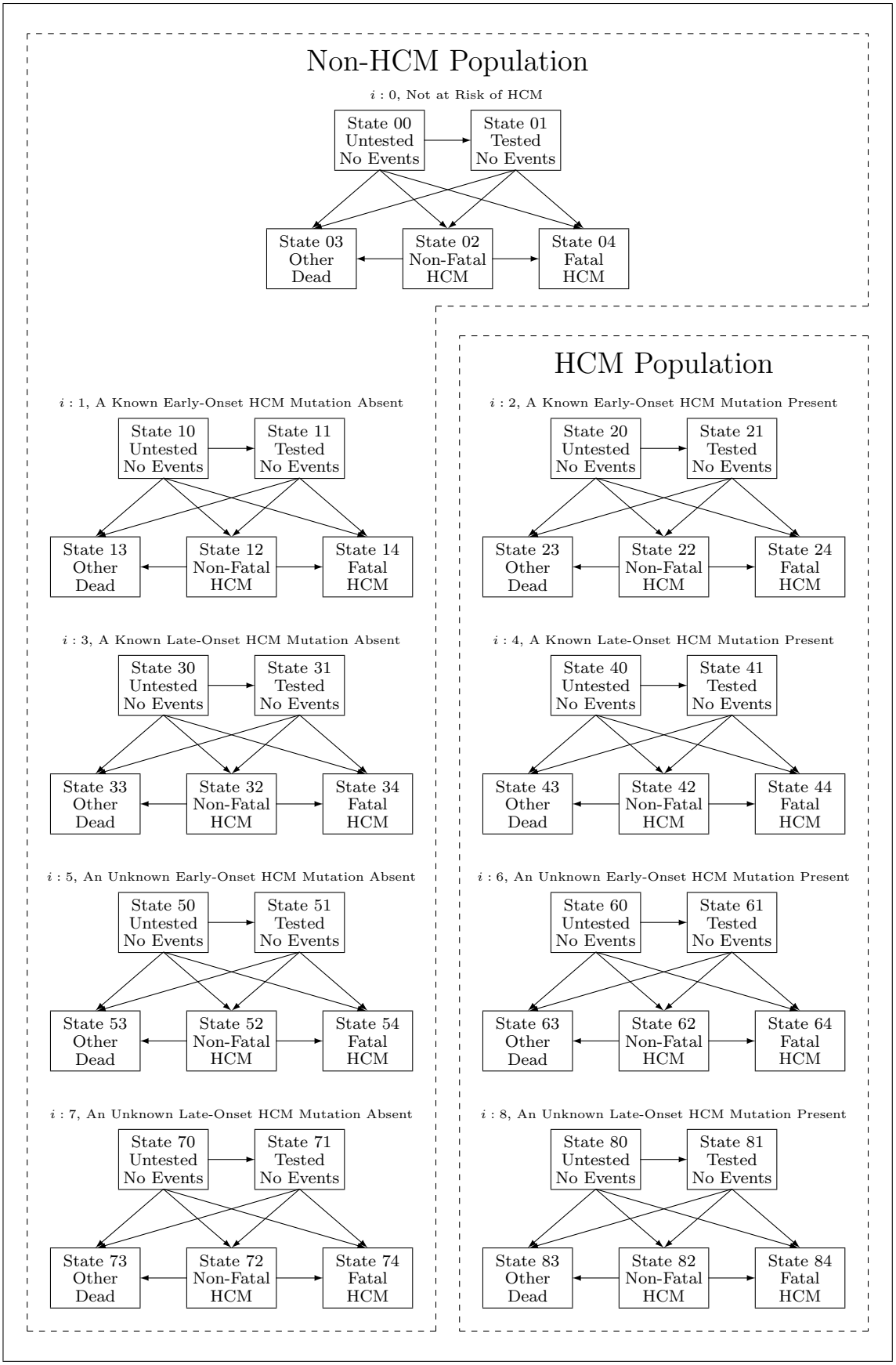


Figure 5.2: A mathematical model of uptake of genetic testing in HCM for a population with nine sub-populations associated with HCM genotype in which each sub-population is classified as being a part of the HCM or non-HCM population.

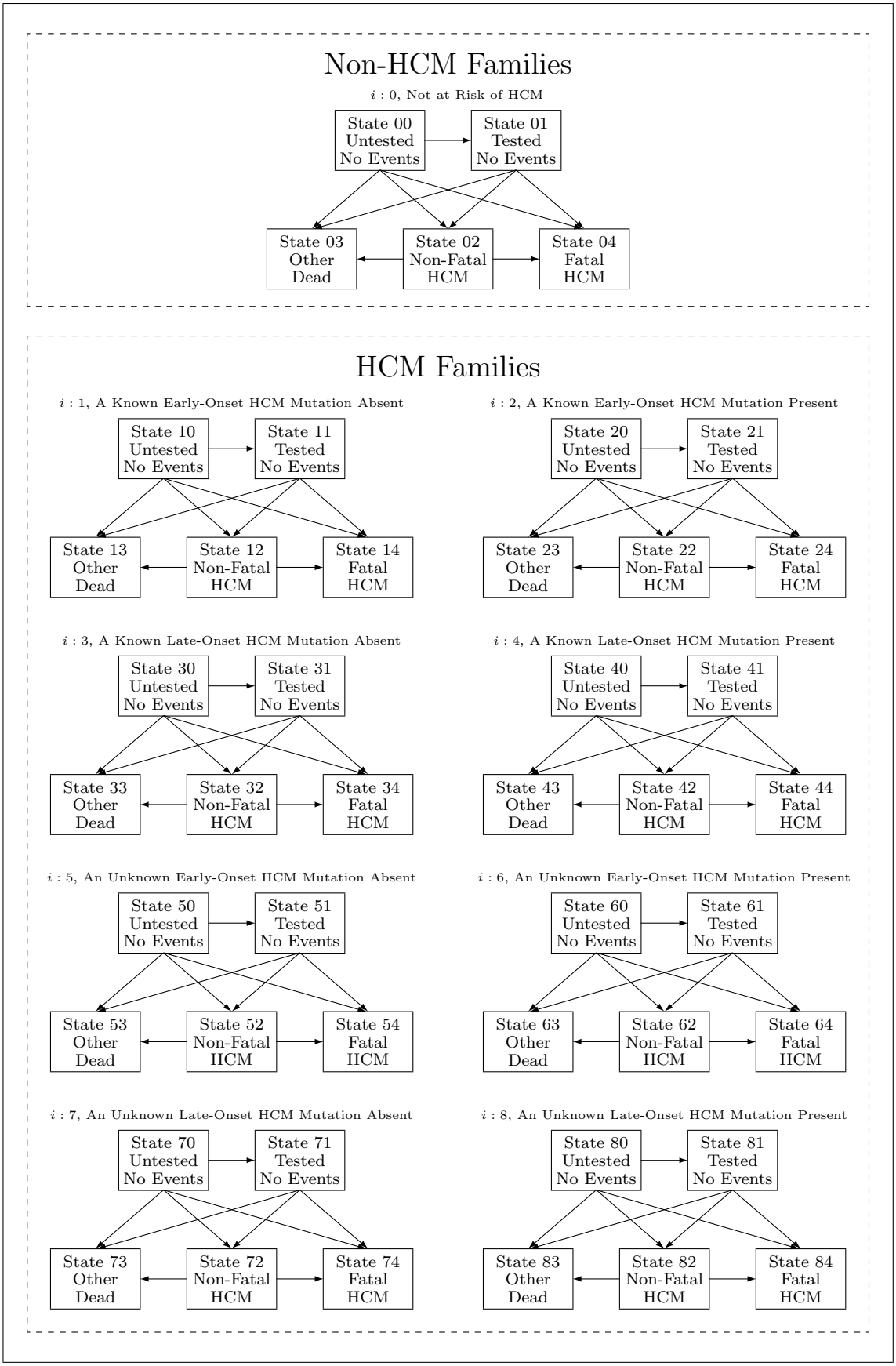


Figure 5.3: A mathematical model of uptake of genetic testing in HCM for a population with nine sub-populations associated with HCM genotype in which each sub-population is classified as being a part of HCM or non-HCM families.

5.2.1 The Necessity of Simulation in the Testing Model

Occupancy probabilities in a multiple-state Markov model can be found by solving the Kolmogorov forward equations (see Section 3.4) with the help of a numerical method such as the fourth-order Runge-Kutta method (see Section 3.5). We obtain occupancy probabilities in the epidemiological model in this way.

However, these techniques cannot be applied to obtain occupancy probabilities in HCM families in the testing model due to the nature of cascade genetic testing in these families (see Section 4.2). Briefly, in cascade genetic testing, we need a proband (index patient), who will trigger genetic testing in some or all of the family members or relatives in a cascade style. It means that until a proband exists in a family, the uptake rate of testing among all family members is zero.

The appearance of a proband in a family is an event that occurs at a random time and which changes the subsequent testing behaviour of at-risk relatives. We do not know when a proband may appear, nor who it will be. For that reason, we develop a stochastic simulation method to simulate cascade genetic testing in HCM families. We mainly aim to simulate insurance cash flows directly in each simulated sample path of the model, which will be discussed in Chapter 7. However, in this chapter, in Section 5.9.3, we present simulated life histories and occupancy probabilities in the testing model states in respect of HCM families during the whole life time.

5.2.2 The Parameters of the Testing Model

We need prevalence rates of HCM-related mutations, which were reported in Section 3.7. Also, we need the transition intensities which are the key parameters for the simulation model. Many of them were reported in detail in Sections 3.7, 3.8, 3.9 and 3.10 in Chapter 3 where we modelled the epidemiology of HCM. Table 3.10 summarized baseline assumptions for the epidemiological model. We adopt these baseline assumptions for the testing model in this chapter and the rest of the study. We also include the baseline uptake rate of genetic testing in Section 4.3.2. And, we present our baseline assumptions for the testing model in Table 5.1. These are used for all the computations in the following sections. Note that the average number of

Table 5.1: Baseline assumptions for the testing model parameters.

Epidemiological Parameters	Table 3.10	Section 3.11
Prevalence of non-HCM mutations in the general population at age 20	0.998	Section 3.7.1
Prevalence of HCM mutations in the general population at age 20	0.002	Section 3.7.1
Prevalence of known early-onset mutations in the HCM population at birth	0.5	Section 3.7.2
Prevalence of known late-onset mutations in the HCM population at birth	0.1667	Section 3.7.2
Prevalence of unknown early-onset mutations in the HCM population at birth	0.25	Section 3.7.2
Prevalence of unknown late-onset mutations in the HCM population at birth	0.0833	Section 3.7.2
Penetrance of early-onset HCM at age 20	100%	Section 3.8.1
Penetrance of late-onset HCM at ages 20–70	Figure 3.4	Section 3.8.2
Hazard rate of fatal HCM per annum for all ages	0.0055	Section 3.9.7
Hazard rate of non-fatal HCM per annum for all ages	Table 3.6	Section 3.9.7
Hazard rate of all other death per annum for all ages	Figure 3.9	Section 3.10
Hazard rate of testing in one year at ages 0–70 if proband exists in family	0.6931472	Section 4.3.2

* See also the parametrisation of average number of children per each family in Section 5.6.1.

children per each family has not been described yet. Therefore, this parameter is not shown in Table 5.1, but it will discussed and parametrised later in Section 5.6.1.

5.2.3 Time Steps in the Numerical Computations

The Kolmogorov forward equations, when needed, in this chapter are numerically solved with time step 0.0005 years (see Section 3.11). We use time step 0.005 years for the simulation of the life histories of HCM families, which we find to be practically as accurate as using 0.0005 years. See the supplementary figures in Section 8.6.3.

5.3 Creation of Families

5.3.1 Nuclear Families

In the testing model, the general population is treated as a population consisting of independent nuclear families. A nuclear family is a family including both parents and their children. We allow the number of children to be zero. See Section 5.1.

A nuclear family embodies two generations: the generations of parents and children which we label respectively as the zeroth and first generations. In order to model, simultaneously, the life histories of all family members, the natural time scale to adopt is calendar time. At any given calendar time, different family members will be of different ages. For simplicity, it is assumed that individuals in the same generation are always born at the same calendar time. We assume all children are born when their parents are 30 years old, so the calendar time at birth for the zeroth generation is 0 while that of the first generation is 30.

5.3.2 Critical Times

The critical times in our stylized model of a family are as follows:

(a) **Calendar time $t = 0$:** Persons of the zeroth generation are born.

(b) **Calendar time $t = 20$:** Persons of the zeroth generation (if still alive) do two things. Firstly, they pair off randomly in male and female couples to

form families. Secondly, they enter the insurance market, and may move from 'uninsured' to 'insured' states (see Chapter 7).

(c) **Calendar time $t = 30$:** Persons of the first generation are born to surviving couples of the zeroth generation.

(d) **Calendar time $t = 50$:** Persons of the first generation enter the insurance market at age 20.

(e) **Calendar time $t = 60$:** Persons of the zeroth generation exit the insurance market: purchase of new insurance policies ceases.

(f) **Calendar time $t = 90$:** Persons of the first generation reach age 60 and exit the insurance market.

We expand on some of the details of this process below.

5.3.3 Prevalence Rates at Calendar Time $t = 0$

The prevalence rate of HCM-related mutations is about 0.002 at ages 23–35 in the general population (Maron et al. 1995), which is conservatively assumed to apply at calendar time and age 20 in the zeroth generation in our model (Table 5.1). (Note that Maron et al. (1995) was based on observation of clinical HCM rather than genetic testing, but we still adopt this rate as a baseline in our model because it is conservative for our purposes, see Section 3.7.1. Also, Howard (2014) relied on the same rate to describe the prevalence of HCM-related mutation carriers in the general population, see Section 1.3.)

However, we need the prevalence rate of HCM-related mutation carriers at calendar time and age zero in the zeroth generation. Therefore, we calculate occupancy probabilities in alive states at age and calendar time 20 in the zeroth generation in the mutation carrier sub-populations ($i = 2$, $i = 4$, $i = 6$, and $i = 8$) in HCM population. Note that until this time, the Kolmogorov forward equations can still be used even when we consider families because there is no family established such that the appearance of a proband can affect the testing behaviour of at-risk relatives. In

Table 5.2: The prevalence rate in non-HCM and HCM populations in the zeroth generation for males and females separately. See Table 5.1.

Population	Gender	Calendar Time and Age 20	Calendar Time and Age 0
Non-HCM	Male	0.998	≈0.99774
	Female	0.998	≈0.99774
HCM	Male	0.002	≈0.00226
	Female	0.002	≈0.00226

Table 5.3: The prevalence rate in each sub-population in the zeroth generation for males and females separately. See Table 5.1.

Calender Time and Age 0				
Population	Gender	Sub-population	State ij	Prevalence Rate
Non-HCM	Male	$i = 0$	00	0.99774
	Female	$i = 0$	00	0.99774
HCM	Male	$i = 2$	20	0.00113
	Female	$i = 2$	20	0.00113
	Male	$i = 4$	40	0.000377
	Female	$i = 4$	40	0.000377
	Male	$i = 6$	60	0.000565
	Female	$i = 6$	60	0.000565
	Male	$i = 8$	80	0.000188
	Female	$i = 8$	80	0.000188

other words, up to calendar time $t = 20$, the life history of a single individual in the testing model is Markov.

Recall that ${}_sp_x^{ijk}$ (see equation (3.2)) denotes the probability that a person who is in state ij at age x will be in state ik at age $x + s$, for males and females separately. The baseline prevalence rates of mutation carrier sub-populations ($i = 2$, $i = 4$, $i = 6$, and $i = 8$) in HCM population, at calendar time and age zero in the zeroth generation for males and females separately can be calculated as follows (see Tables 5.2 and 5.3):

$$\frac{0.002}{{}_{20}p_0^{200} + {}_{20}p_0^{400} + {}_{20}p_0^{600} + {}_{20}p_0^{800}}, \tag{5.1}$$

see also equation (3.31).

Up to calendar time 20, we treat the general population as five sub-populations: four HCM mutation carrier sub-populations ($i = 2, 4, 6, 8$) with initial prevalence rates, shown in Table 5.3; and a single non-carrier sub-population ($i = 0$) with initial

prevalence $(1 - 0.00226) = 0.99774$. We keep males and females separate.

Then, at calendar time 20, we model family formation by removing individuals who "marry" HCM mutation carriers (in sub-populations $i = 2, 4, 6, 8$) from sub-population $i = 0$ and placing them in sub-populations $i = 1, 3, 5, 7$ respectively. Thus sub-populations $i = 1, 3, 5, 7$ are empty until calendar time 20, when they are populated by the spouses of mutation carriers.

5.3.4 Population Size and Composition at Calendar Time $t = 0$

Start at calendar time $t = 0$ with a fixed population of 5 million persons (2.5 million each of males and females) at age zero, which is sufficient to satisfy the actuarial equivalence principle of insurance losses under no adverse selection in our model, see Chapters 6 and 8.

The number of each sex in the HCM population (the collection of four carrier sub-populations $i = 2$, $i = 4$, $i = 6$, and $i = 8$) is:

$$2,500,000 \times 0.00226 = 5,650, \tag{5.2}$$

and in the non-HCM population (non-carrier sub-population $i = 0$):

$$2,500,000 \times 0.99774 = 2,494,350. \tag{5.3}$$

5.4 Family Formation: HCM Families

Table 5.4 shows the number alive at calendar time (and age) 20 in each carrier sub-population, to each of whom a spouse has been allocated to form a family. We explain by means of an example. Note that the numbers below are all rounded to integer values.

- At calendar time $t = 20$, there are $2,467,999$ surviving males in the non-carrier sub-population (state 00) and $2,476,010$ surviving females.

Table 5.4: Total number of HCM families consisted of alive carriers and their assigned alive non-carrier spouses at calendar time and age 20.

Carrier Population			Spouse Population		
Gender	State ij	Size	Gender	State ij	Size
Male	20	2406	Female	10	2406
Female	20	2413	Male	10	2413
Male	40	932	Female	30	932
Female	40	935	Male	30	935
Male	60	1203	Female	50	1203
Female	60	1207	Male	50	1207
Male	80	466	Female	70	466
Female	80	467	Male	70	467
Male	22	237	Female	10	237
Female	22	238	Male	10	238
Male	42	0	Female	30	0
Female	42	0	Male	30	0
Male	62	119	Female	50	119
Female	62	119	Male	50	119
Male	82	0	Female	70	0
Female	82	0	Male	70	0
		10,742			10,742

- At calendar time $t = 20$, there are 2406 surviving males in the 'known early-onset mutation carrier' sub-population who have not suffered an HCM event (state 20) and 2413 surviving females.

- At calendar time $t = 20$, each of those 2406 surviving males is allocated a female spouse who moves from state 00 to state 10. Hence state 10 (females) now contains 2406 persons, and the number in state 00 (females) is 2406 less.

- We do the same for female survivors in state 20, thus populating state 10 (males) with 2413 individuals and reducing the number in state 00 (males) by the same number.

- We do the same for the 237 male survivors and 238 female survivors in state 22 (males and females respectively) thus adding a further 237 and 238 individuals to each of state 10 (females and males) and reducing the numbers in state 00 (males and females) accordingly.

- We repeat the process for all the other carrier sub-populations ($i = 4, 6, 8$) and

then again for individuals in each carrier sub-population who have survived but suffered an HCM-related event (states 42, 62, 82).

- At the end of this process the non-carriers' state 00 contains $2,462,620$ males and $2,470,647$ females.

Note that we ignore the small probability that both spouses are mutation carriers. All HCM families at calendar time $t = 20$ consists of one carrier and a non-carrier spouse.

5.5 Family Formation: Non-HCM Families

The individuals in non-HCM families are not affected with any HCM-related mutations at all. No one in these families will ever have a genetic test, and (in Chapter 7) their insurance purchasing behaviour remains constant. Therefore, numerical solution of the Kolmogorov forward equations can still be used throughout their lifetime, a great computational advantage.

From the non-carriers remaining in sub-population $i = 0$ (male and female), we can form a total of $2,462,620$ spouse pairs. In due course, at calendar time $t = 30$, children will be born in those families in which both spouses are alive, the same as in HCM families. These families then provide the 'normal' pool of insurance purchasers, against which the cost of adverse selection can be measured (see Chapter 7).

5.6 Population Size and Composition at Calendar Time $t = 30$: HCM Families

5.6.1 Birth of Children

If both spouses are alive at age and calendar time 30, then they have a random number of children. We assume the number of children to have a Poisson (λ) distribution. Since we rely on the recent data for either fatal HCM or all other causes

Table 5.5: Average number of children per female in the US (Martin et al. 2018) and the UK (Office for National Statistics 2018) between years 2013–2017.

Year	US	UK
2013	1.86	1.85
2014	1.86	1.83
2015	1.84	1.82
2016	1.82	1.81
2017	1.77	1.76

of mortality from the US, it would be realistic to look into the average number of children per female in the US. For this purpose, in Table 5.5, we present the average number of children per female in the US (Martin et al. 2018) and compare them with figures for the UK (Office for National Statistics 2018). We assume $\lambda = 1.8$ as a baseline assumption for the average number of children per female (or family) in our model which seems appropriate for both the US and the UK.

5.6.2 Gender of Children

The gender of parents, or individuals in the zeroth generation, is deterministically determined at birth. See Section 5.3. The gender of any child, an individual in the first generation, in an HCM family is randomly determined at birth, being male or female with probability 0.5. Note that the percentage of male versus female babies is not quite 50% in most countries.

5.6.3 Sub-populations

See Sections 5.3 and 5.4 for the sub-populations of carrier parents and their non-carrier spouses in the zeroth generation in HCM families. Each child in an HCM family is randomly assigned a genotype based on Mendel's law. Therefore, each child is allocated to a sub-population at birth. If the carrier parent's sub-population is $i = 2$ (respectively $i = 4$, $i = 6$, $i = 8$), the child is allocated to sub-population $i = 2$ (respectively $i = 4$, $i = 6$, $i = 8$) with probability 0.5, or else to sub-population $i = 1$ (respectively $i = 3$, $i = 5$, $i = 7$).

5.7 Population Size and Composition at Calendar Time $t = 30$: Non-HCM Families

5.7.1 Birth and Gender of Children

We create children for the non-HCM families at calendar time $t = 30$ if both spouses are then alive.

(a) At calendar time $t = 20$, we created $2,462,620$ spouse-pairs, non-HCM families, in sub-population $i = 0$ (Sections 5.4 and 5.5).

(b) At calendar time $t = 30$, we obtain the total number of families in which both spouses survive (rounded to integer values), where we assume the future lifetimes of both spouses are independent, as follows:

$$2,462,620 \times {}_{10}p_{20}^{000,\text{male}} \times {}_{10}p_{20}^{000,\text{female}} = 2,418,869. \tag{5.4}$$

The number and gender of children in non-HCM families is modelled deterministically at calendar time $t = 30$, we assume that each family has $\lambda/2$ male children and $\lambda/2$ female children, where λ is as in Section 5.6.1. Since $\lambda = 1.8$ is our baseline assumption, the numbers of male and female children separately at age zero at calendar time $t = 30$ are (rounded to integer values):

$$2,418,869 \times 0.9 = 2,176,982, \tag{5.5}$$

which we will refer to as the first generation in non-HCM families.

5.7.2 Sub-populations

These children in non-HCM families will also be in sub-population $i = 0$ at birth. From this point on, family relationships do not matter in the $i = 0$ sub-population, since there is no genetic testing and no adverse selection. We do not need to resort to simulation of these life histories, we can calculate occupancy probabilities

directly using the Kolmogorov equations and (in Chapter 7) expected present value of insurance cash flows using Thiele's equations.

5.8 Non-HCM Families at Calendar Times $t = 20 - 90$

At calendar times 20–90, the life histories of parents and children in non-HCM families, or in sub-population $i = 0$, are always Markov, since they never have genetic testing. Therefore we can always find occupancy probabilities in any state in this sub-population by solving the Kolmogorov forward equations. Individuals in these families, or in sub-population $i = 0$, will no longer be discussed in this chapter because they are never genetically tested. Nevertheless, they will be vital when we estimate adverse selection costs because they will be 'normal' purchasers of insurance.

5.9 HCM Families at Calendar Times $t = 20 - 90$

5.9.1 Simulating Life Histories of HCM Families

An HCM family consists of one mutation carrier parent, one non-carrier spouse and a number of children (which can be zero), each a carrier or non-carrier according to Mendel's law. Each member is identified by generation, age, gender, and sub-population. We assume all families to be statistically independent of each other.

We now describe our simulation algorithm, which simulates simultaneously in calendar time each family member's life history over short time steps. This process allows the transition intensities in respect of each person to depend on the life histories of all family members, see Section 5.1.

In what follows, we give the recipe of the simulation algorithm, which was programmed in C++. The algorithm simulates the transitions made from calendar time t to $t + dt$ by each family member simultaneously over a suitably small time step of length dt.

(a) Suppose the family has γ members, labelled by $1, 2, ..., \gamma$.

(b) Loop through the γ family members one by one.

(c) Suppose the rth member ($r = 1, 2, ..., \gamma$) is in state ij (that is sub-state j of sub-population i) and at calendar time t is age x.

(d) Approximate the probability of moving from state ij to state ik in time dt, from equation (3.4), by:

$$ {}_{dt}p_x^{ijk} \approx \mu_x^{ijk} dt, \quad j \neq k. \tag{5.6}$$

Note that if $X(t)$ is the occupied state at calendar time t, then ${}_{dt}p_x^{ijk}$ is:

$$ {}_{dt}p_x^{ijk} \approx P\left[X(t+dt) = ik | X(t) = ij\right]. \tag{5.7}$$

(e) Check that $\sum_k \mu_x^{ijk} dt < 1$, ($dt$ has to be small enough that this check never fails) and simulate a U(0,1) random variable, denoted by $\mathcal{U}$.

(f) If $\sum_{l=0}^{k-1} \mu_x^{ijl} dt < \mathcal{U} \leq \sum_{l=0}^{k} \mu_x^{ijl} dt$ ($k = \{1, 2, 3, 4\}$) record transition from state ij to state ik during time dt. Otherwise, record that the family member did not leave state ij during time dt.

5.9.2 Genetic Testing Behaviour in HCM Families

We now consider the genetic testing behaviour of individuals, respectively, in different sub-populations.

(a) Figure 5.4 represents individuals in sub-population $i = 0$. They never have genetic tests.

(b) Figure 5.5 and Figure 5.7 represent non-carrier and carrier individuals, respectively, in families affected with a known HCM mutation. It is assumed that nobody would be tested in an HCM family if no proband exists in the family. If a proband exists in the family, then the testing behaviour of at-risk relatives

changes. If the proband is the carrier parent, the spouse of the carrier parent would not take up genetic testing, as their risk of carrying a mutation is unchanged, but the children of the carrier parent would take up genetic testing at some assumed rate. If the proband is a carrier child, the siblings of the carrier child and both parents would take up genetic testing at some assumed rate. Their insurance purchasing behaviour might change either because of test results if they were tested or because of their Mendelian risk if they were not tested, see Chapter 7. Note that we assume the testing is a joint decision in both parents.

(c) Figure 5.6 and Figure 5.8 represent non-carrier and carrier individuals, respectively, in families affected with an unknown HCM mutation. In this case, it is assumed that the proband is tested and found not to carry any known HCM mutation. Therefore, at-risk family members would be monitored clinically but not genetically tested. Their insurance purchasing behaviour might change because of their Mendelian risk, see Chapter 7.

5.9.3 Simulation Results

In Section 5.9.1, we explained how we simulate life histories in HCM families. In this section, we show the mean and standard deviation of 500 independent simulations of the occupancy probabilities in the testing model states in respect of HCM families.

In Figure 5.9, we present the mean of the occupancy probabilities of simulated lives of the parents/the zeroth generation obtained from 500 independent simulations in each state in the testing model in respect of the HCM families. In Figure 5.9:

- Females (compare to males) are more likely to occupy state 0, untested no events (the upper left plot) at calendar time 20–90 because they are less likely to be exposed to the risks of late-onset HCM and have lower all-cause mortality than males. See Sections 3.8.2 and 3.10.

- No parent takes up genetic testing at calendar time 20–30 (state 1, tested no events, the upper right plot) because we form children to families at calendar

time 30. That is why, there would be no child probands during this time that can make their parents take up genetic testing.

In Figure 5.10, we present the mean of the occupancy probabilities of simulated lives of the children/the first generation obtained from 500 independent simulations in each state in the testing model in respect of the HCM families. In Figure 5.10:

- Just after age zero and calendar time 30 in the first generation, there is a jump, significant decrease, in the occupancy probabilities in state 0, untested no events (the upper left plot) due to the infant mortality (see Figure 3.9). For the same reason, there is a jump in the occupancy probabilities in state 3, other dead (the lower left plot).

- Just after age zero and calendar time 30 in the first generation, there is a jump in the occupancy probabilities in state 1, tested no events (the upper right plot) because we conservatively (for insurance purposes) assume children can take up genetic testing at birth if they have a proband parent (see Section 4.3.2).

The standard deviations of Figures 5.9 and 5.10 are observed as approximately zero at all ages (due to very large number of individuals in the simulation), so they will not be displayed.

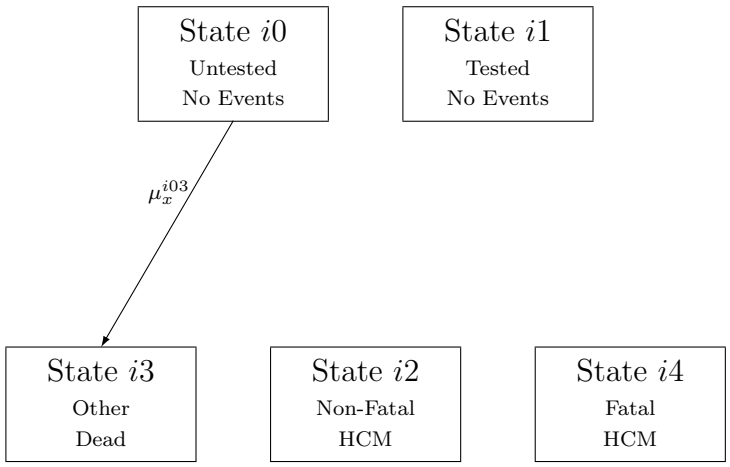
5.10 Discussion

For the same reason as in Section 3.11, we do not make any sensitivity analysis for the results reported through this chapter because our study mainly focuses on measuring the insurance costs under adverse selection. Therefore, the sensitivity assumptions for the model parameters, see Chapters 3 and 4, will appear in the sensitivity analysis for the adverse selection costs in Chapter 9.

Furthermore, during this chapter, we established a mathematical model of cascade genetic testing in HCM, called 'the testing model'. As with our epidemiological model in Chapter 3, this model also can be treated as a medical model. However,

some may argue whether or not the testing model answers the questions that clinical practitioners are interested in rather than insurers. In daily practice, if a proband arises in an HCM family, clinicians recommend the proband to see a geneticist to learn the genetic substrate of the disorder and let the relatives know about their risks to carry the identical mutation. The questions mostly asked by the geneticist are about how they would manage the risks of developing the disorder (onset of HCM) at different ages. For example, the ESC Guidelines (Elliott et al. 2014) and the ACCF/AHA Guidelines (Gersh et al. 2011) recommend a life-long clinical check-up of the heart at regular time intervals for the HCM individuals tested with carrying the HCM-related mutations, but did not clinically develop the disorder. This is a significant burden both physically and economically for these individuals. A medical doctor might be more interested in finding a way of reducing this burden. There are also upcoming studies focusing on these problems. For example, Wordsworth et al. (2010) presented a cost-effectiveness model which observes slight increases in the life expectancies of the individuals who undergo cascade genetic testing in HCM. Note that the HCM-related hazard rates do not depend on genetic testing in our model.

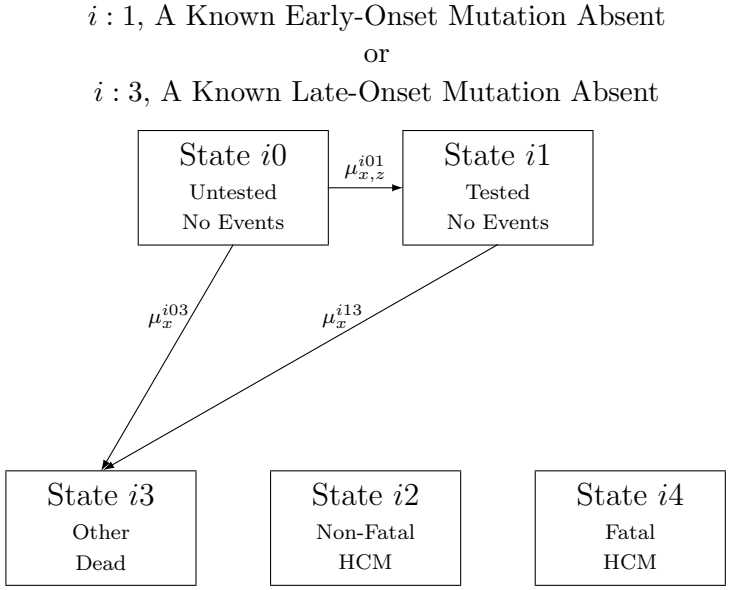
We can now introduce 'a life insurance model of HCM' to measure the impact of adverse selection among individuals with genetic tests, since we have a clear picture of the epidemiology of HCM and of genetic testing in HCM with their corresponding mathematical models in Chapter 3 and Chapter 5. Before doing so in Chapter 7, we discuss the fundamental theory of life insurance mathematics in Chapter 6.



Fixed Rates:

- $\boldsymbol{\mu_x^{i03}}$ = annual mortality rate of all other causes at all ages.

Figure 5.4: A mathematical model of a life history of an individual r, a member of a family in the $i = 0$ risk sub-population in the testing model.



Fixed Rates:

- $\boldsymbol{\mu^{ij3}_{x}}$ = annual mortality rate of all other causes at all ages.

If no proband exists in family:

- $\boldsymbol{\mu^{i01}_{x,z}} = 0$, uptake rate of testing per annum at any age.

If carrier parent becomes proband with a known mutation in family:

(a) *r is a spouse of carrier parent;*
- $\boldsymbol{\mu^{i01}_{x,z}} = 0$, uptake rate of testing per annum at any age.

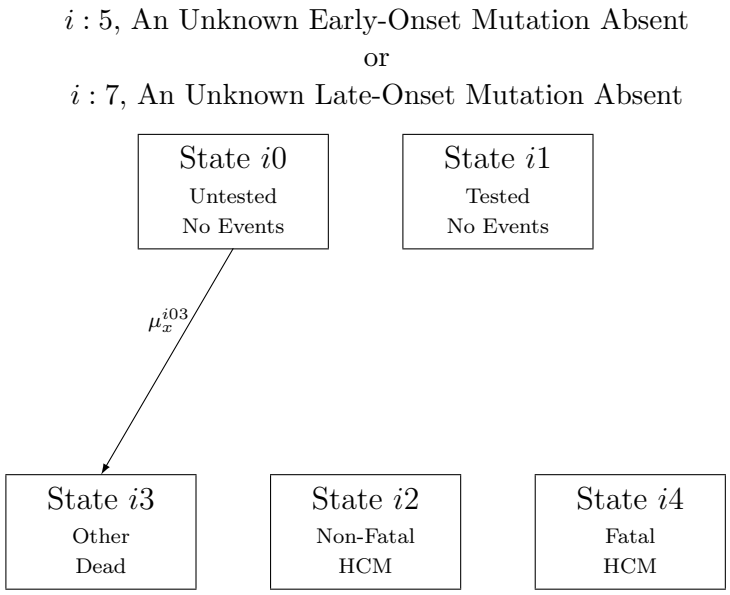
(b) *r is a non-carrier child of carrier parent;*
- $\boldsymbol{\mu^{i01}_{x,z}}$ = normal uptake rate of testing per annum at all ages.

If a carrier child becomes proband with a known mutation in family:

(a) *r is a spouse of carrier parent not tested nor become a subsequent proband; or, r is a non-carrier sibling of the carrier child;*
- $\boldsymbol{\mu^{i01}_{x,z}}$ = normal uptake rate of testing per annum at all ages.

(b) *r is a spouse of carrier parent tested or become a subsequent proband;*
- $\boldsymbol{\mu^{i01}_{x,z}} = 0$, uptake rate of testing per annum at any age.

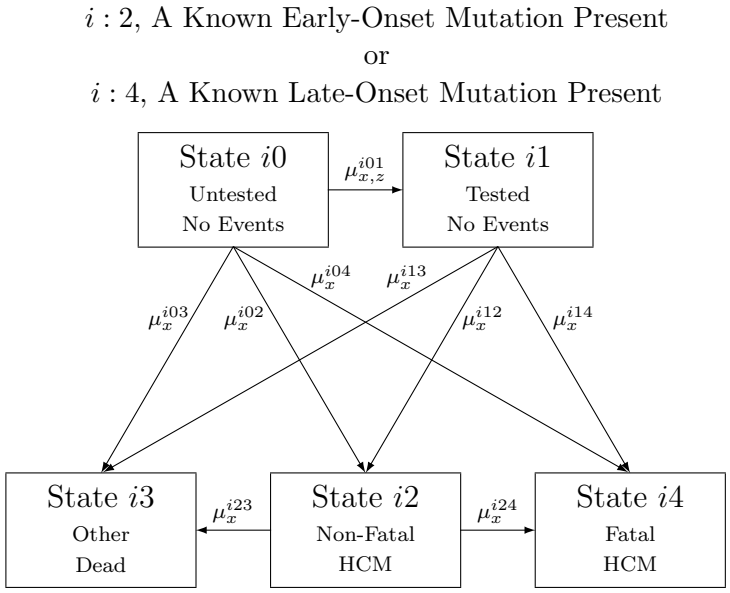
Figure 5.5: A mathematical model of a life history of an individual r, a non-carrier member in which one carrier parent carries a known HCM mutation, in the $i = 1$ or $i = 3$ risk sub-populations in the testing model. In $\mu^{i01}_{x,z}$, z refers to duration in state $i0$ since (if) a proband exists in the family.



Fixed Rates:

- $\boldsymbol{\mu_x^{i03}}$ = annual mortality rate of all other causes at all ages.

Figure 5.6: A mathematical model of a life history of an individual r, a non-carrier member in which one carrier parent carries an unknown HCM mutation, in the $i = 5$ or $i = 7$ risk sub-populations in the testing model.



Fixed Rates:

- $\boldsymbol{\mu^{i02}_x = \mu^{i12}_x}$ = proportioned, respective to penetrance of clinical HCM, non-fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu^{ij3}_x}$ = annual mortality rate of all other causes at all ages.
- $\boldsymbol{\mu^{i04}_x = \mu^{i14}_{x+t}}$ = proportioned, respective to penetrance of clinical HCM, fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu^{i24}_x}$ = fatal HCM rate per annum at all ages.

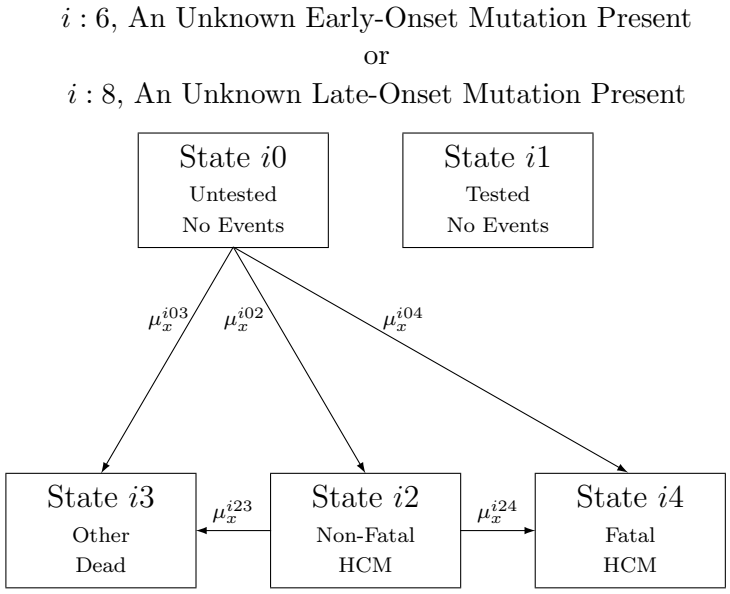
If no proband exists in family:

- $\boldsymbol{\mu^{i01}_{x,z}} = 0$, uptake rate of testing per annum at any age.

If a proband exists (who not matter) with a known mutation in family:

- $\boldsymbol{\mu^{i01}_{x,z}}$ = normal uptake rate of testing per annum at all ages.

Figure 5.7: A mathematical model of a life history of an individual r, a carrier member in which one carrier parent carries a known HCM mutation, in the $i = 2$ or $i = 4$ risk sub-populations in the testing model. In $\mu^{i01}_{x,z}$, z refers to duration in state $i0$ since (if) a proband exists in the family.



Fixed Rates:

- $\boldsymbol{\mu_x^{i02}}$ = proportioned, respective to penetrance of clinical HCM, non-fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{ij3}}$ = annual mortality rate of all other causes at all ages.
- $\boldsymbol{\mu_x^{i04}}$ = proportioned, respective to penetrance of clinical HCM, fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{i24}}$ = fatal HCM rate per annum at all ages.

Figure 5.8: A mathematical model of a life history of an individual r, a carrier member in which one carrier parent carries with an unknown HCM mutation, in the $i = 6$ or $i = 8$ risk sub-populations in the testing model.

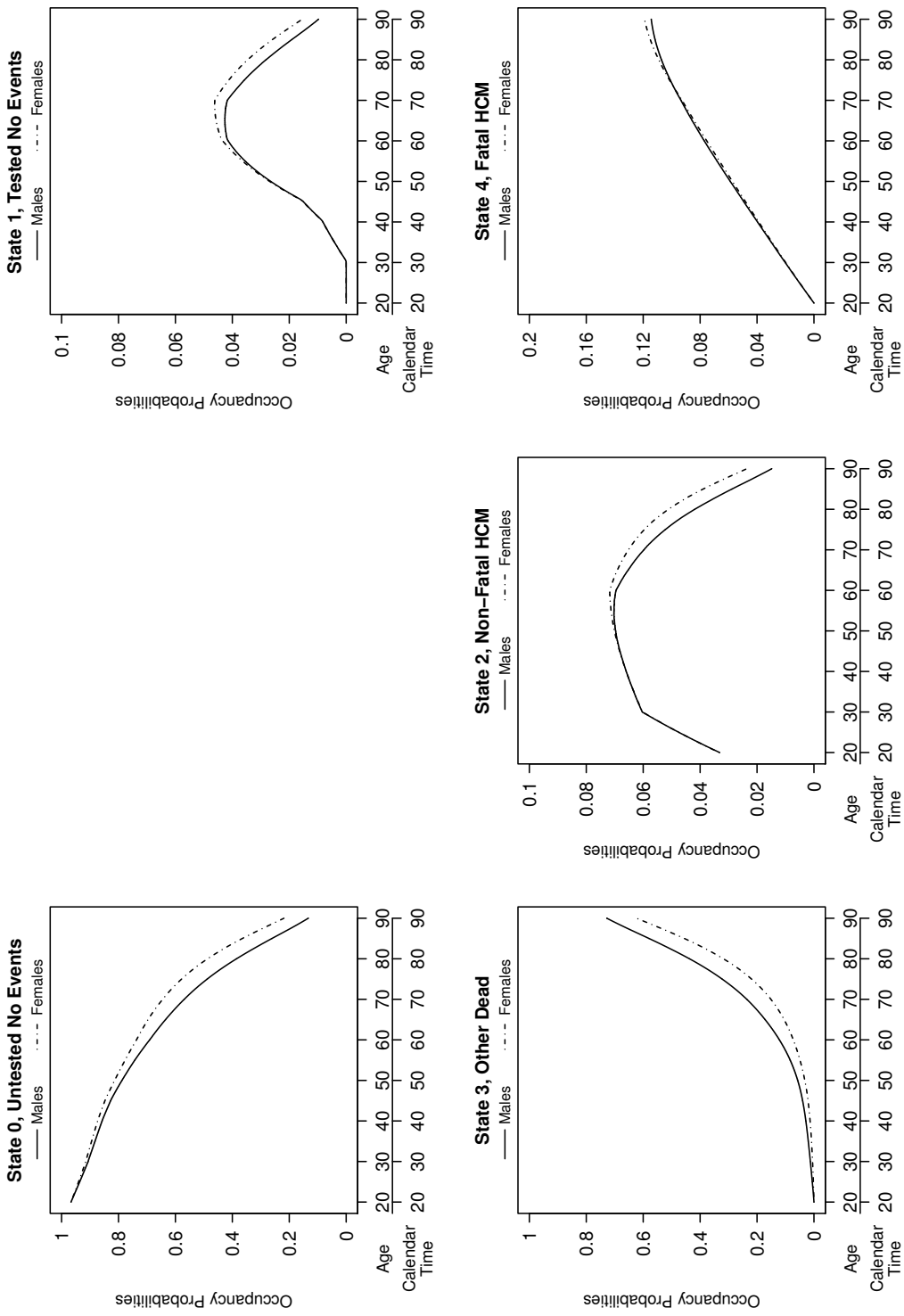


Figure 5.9: Mean of occupancy probabilities in each state in the testing model under baseline assumptions obtained from 500 simulations in the zeroth generation in the HCM families. Note that the vertical axes have different scales.

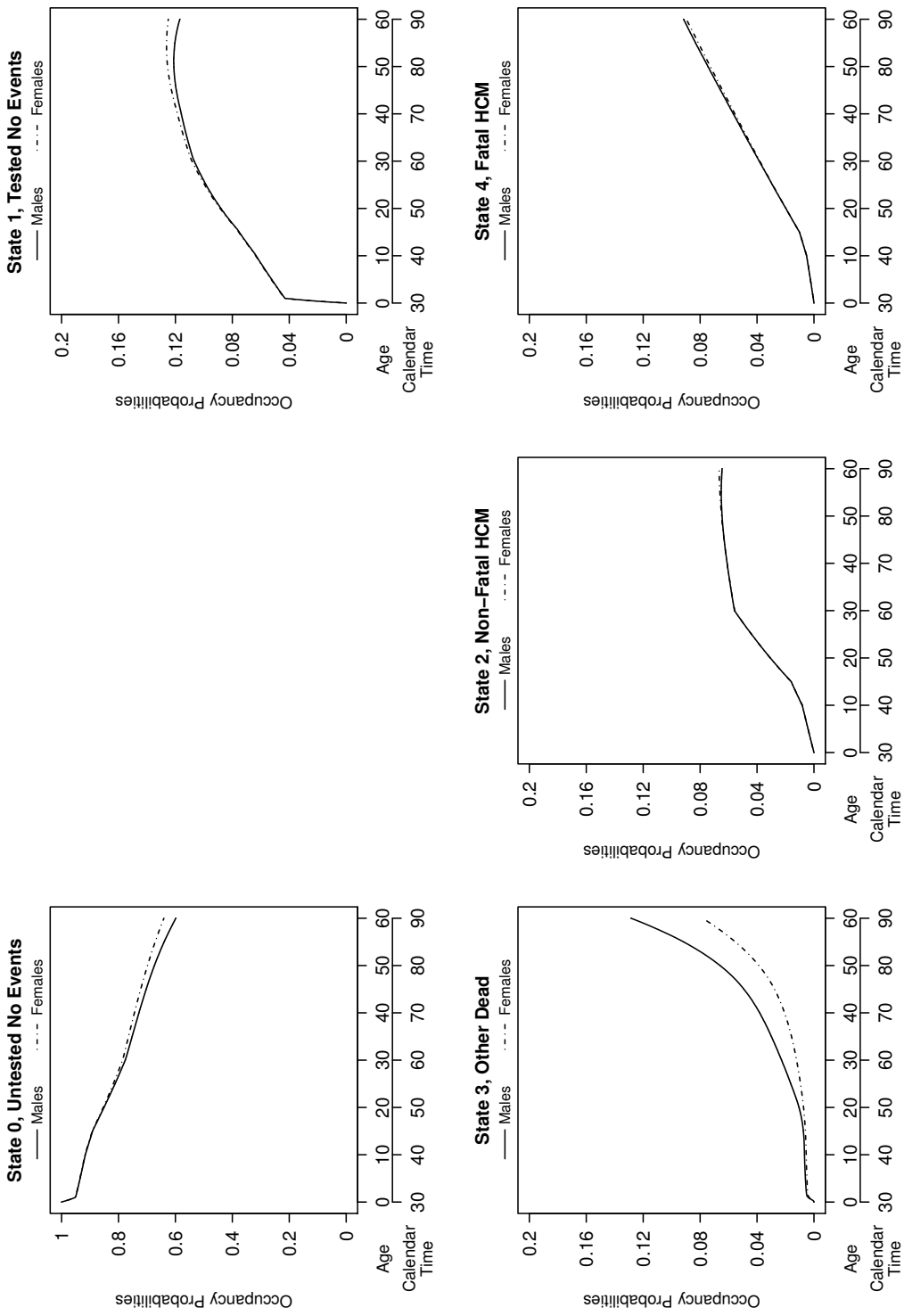


Figure 5.10: Mean of occupancy probabilities in each state in the testing model under baseline assumptions obtained from 500 simulations in the first generation in the HCM families. Note that the vertical axes have different scales.

Chapter 6

Life Insurance Mathematics

6.1 Introduction

In this chapter, we discuss principles of life insurance mathematics. So far, we have developed two models: an epidemiological model of HCM in Chapter 3 and a simulation model of cascade genetic testing in HCM in Chapter 5. Both models focused on the life histories of individuals with/without the risk of HCM. The epidemiological model, in Chapter 3, was a multiple-state Markov model, which models the life history of a single individual, a carrier or a non-carrier of an HCM mutation. The critical endpoint in the epidemiological model was fatal HCM, to the risk of which HCM mutation carriers only are exposed. The testing model, in Chapter 5, was also a multiple-state model, similar to the epidemiological model with more model states. However, in the testing model, the life history of a single individual was not always Markov any more because transition intensities representing uptake of genetic testing by individuals depended on family history as well as on the currently occupied state. The transition intensity from an 'untested' state to a 'tested' state also depended on the duration since a family history (proband) appeared.

The adverse selection model, in Chapter 7, is based on the testing model, extended with the addition of 'insured' states. In addition to simulating the life histories of family members, it will record the cash flows associated with the purchase of life insurance. From the adverse selection model, we will present our measure of the adverse selection costs in Chapter 8 and obtain the results with regard to the

measure in Chapter 9.

In this chapter, we introduce the necessary life insurance mathematics that will be fundamental in Chapters 7, 8, and 9. In Section 6.2, we give a very basic introduction to life insurance mathematics. In Section 6.3, we model cash flows and insurance losses. In Section 6.4, we model insurance losses based on the life history of a single individual, represented by a two-state model. In Section 6.5, we extend the two-state model of insurance losses into multiple-state models for the life history of a single individual. In Section 6.6, we define policy values. All these sections proceed under the assumption that there is no adverse selection for insurers. In Section 6.7, we evaluate insurance losses and policy values under adverse selection. In Section 6.8, and Appendix A, we consider how to extend the model of a single life in a known population to a model in which a life may be in one of several populations, which one being possibly not known by an insurer or by the individual themselves.

Good references for this chapter are Norberg (1991, 1992), Cairns et al. (1998), Dickson et al. (2013), and Macdonald et al. (2018).

6.2 Basics of Life Insurance Mathematics

A life insurance policy is a financial product, sold by a company called the *insurer*, to an individual life, namely the *insured*, and pays *benefits*, generally a lump sum payment to the beneficiaries of the insured in the event of the death of the insured. The insured pays *premiums* to the insurer which represents the price of the product. The premium payments are commonly made at regular intervals (such as every month, every year, etc.). A series of regular premium payments is technically identified as an *annuity*.

The business agreement between the insurer and the insured is legalized by a contract called an *insurance policy*. The time during which the policy will be in force is called the *policy term*. An insurance policy mainly regulates the following obligations for both parties; the insurer who sells the policy and the insured, also called the *policyholder*, who buys the policy:

(a) The obligation of the insurer is to fulfil the liability of paying a benefit, if the

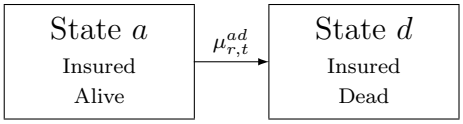


Figure 6.1: A mathematical model of life insurance for insured individual r.

insured dies during the policy term, to the beneficiaries of the insured.

(b) The obligation of the insured is to make a series of regular premium payments, until the event of death or the policy ends, to the insurer.

Life insurance mathematics describes money transactions between the insurer and the insured. By convention, we regard payments by the insurer as positive, and payments to the insurer as negative. We will not discuss all types of life insurance products. We will focus on 'term life insurance', which can also be designed in many ways. In our study, the features of a term life insurance product are that a lump sum payment is paid by the insurer as an outgo, if the insured dies during the policy term, and premiums are paid by the insured as an income, as long as the insured is alive during the policy term or until the policy ends.

In the following section, we construct a mathematical function of the payment stream evaluating cash flows in respect of term life insurance.

6.3 Cash Flows and Insurance Losses

An insurance loss is a random variable which measures the present value of cash flows from an insurance product created by benefit paid and premium income earned, depending on the random event of death of the insured during the policy term. We model the event of the death of the insured as a random event indexed by time, or a stochastic process.

Cairns et al. (1998) formulate loss random variables in the framework of stochastic processes, in particular in a two-state model. We evaluate insurance losses in Section 6.4 for a single individual from such a simple model. We obtain insurance losses in Section 6.5 for a single individual in multiple-state models referring to the

work of Norberg (1991, 1992). We present policy values of these models in Section 6.6. Until Section 6.7, we assume there is no adverse selection for insurers. Section 6.7 presents an actuarial measure of losses suffered by an insurer if an adverse selection risk arises. See also Macdonald et al. (2018) for background of the mortality basis in these sections.

In Chapter 7, our time unit will be calendar time rather than age. It is convenient to define $x_r(t)$ to be the age, at calendar time t, of an individual labelled by r. In this chapter, we define transition intensities of the general form $\mu_{r,t}$ to mean a transition intensity applying to individual r at calendar time t, making use of $x_r(t)$ if the transition intensity depends on age.

6.4 Insurance Losses in a Two-State Model

Figure 6.1 represents the life history of insured individual r in a two-state model labelled by state a for 'insured alive' and state d for 'insured dead' where we suppose the insurance contract was purchased at calendar time zero. We denote by $\mu_{r,t}^{ad}$ the transition intensity of insured individual r per annum between state a and state d at calendar time t, which depends only on calendar time t, and therefore satisfies the Markov assumption (see Assumption 3.1). Thus, Figure 6.1 can also be viewed as a simple case of a Markov model.

In Chapter 3, we gave an explanation of transition intensities in a multiple-state Markov model, which we used to model the epidemiology of HCM. The numerical solution of the Kolmogorov forward differential equations, in Sections 3.4 and 3.5, can be applied to the two-state alive-dead model to obtain occupancy probabilities in both states. More generally, we formulate a stochastic process model of the payment stream associated with a life insurance contract.

6.4.1 Benefit Outgo in the Alive-Dead Model

Benefit payment is made in the event of the death of an insured individual r. Therefore, let $N_r^{ad}(t)$ count the number of transitions of insured individual r from state a to state d up to and including calendar time t. $N_r^{ad}(t)$ is a 0-1 or one-jump counting

process, with non-decreasing and right-continuous sample paths, defined as follows:

$$N_r^{ad}(t) = \begin{cases} 1, \text{ insured individual } r \text{ is dead at calendar time } t; \\ 0, \text{ insured individual } r \text{ is alive at calendar time } t. \end{cases} \tag{6.1}$$

The benefit payment is the lump sum payment of amount $A_r^{ad}(t)$ (assumed to be previsible) if insured individual r dies at calendar time t. In regard to continuous time, we can express the benefit outgo to insured individual r at calendar time t as follows:

$$A_r^{ad}(t) dN_r^{ad}(t) \tag{6.2}$$

where $dN_r^{ad}(t)$ counts the number of transitions of insured individual r from state a to state d between calendar time t and $t + dt$:

$$dN_r^{ad}(t) = \lim_{dt \to 0} \left[N_r^{ad}(t) - N_r^{ad}(t - dt) \right]. \tag{6.3}$$

6.4.2 Premium Income in the Alive-Dead Model

Premium income is obtained as long as insured individual r is alive and the policy is in force. Therefore, define $I_r^a(t)$ to be an indicator function as follows:

$$I_r^a(t) = \begin{cases} 1, \text{ insured individual } r \text{ is alive at calendar time } t^-; \\ 0, \text{ otherwise.} \end{cases} \tag{6.4}$$

Then denote by $a_r^a(t)$ (assumed to be previsible) the annual rate of continuous premium payment made by insured individual r alive at calendar time t^-. In regard to continuous time, we can express the premium income from insured individual r between calendar time t and $t + dt$ as follows:

$$a_r^a(t) I_r^a(t) dt. \tag{6.5}$$

6.4.3 Insurance Loss in the Alive-Dead Model

The insurance loss arising from insured individual r in the alive-dead model between calendar time t and $t + dt$, denoted by $dL_r(t)$, is defined as:

$$dL_r(t) = A_r^{ad}(t)dN_r^{ad}(t) - a_r^a(t)I_r^a(t)dt. \tag{6.6}$$

The cumulative insurance losses from calendar time zero up to calendar time T (assuming payments up to calendar time T where T can be ∞) from insured individual r is:

$$L_r(T) = \int_0^T dL_r(t). \tag{6.7}$$

The present value of the insurance losses at calendar time zero, with the constant force of interest δ per annum, is:

$$L_r = \int_0^T e^{-\delta t} dL_r(t). \tag{6.8}$$

6.4.4 The Actuarial Equivalence Principle

The actuarial equivalence principle allows the premiums should be determined at outset such that

$$E\left[L_r\right] = 0. \tag{6.9}$$

This principle implies that $a_r^a(t)$ can be any function of calendar time t as long as equation (6.9) is satisfied.

(a) Most commonly, the premiums are paid at a constant rate such that $a_r^a(t) = \pi_r$ at calendar time t where π_r denotes the level premiums made by individual r in state a.

(b) However, in our model, the premiums will not be constant for each individual ($a_r^a(t) \neq \pi_r$) because in Chapter 7 et. seq. it will be convenient if all insured individuals pay the same rate of premiums depending on their current age (or the current calendar time), but not the age (or the calendar time) at which they purchased their insurance policy.

(c) Instead, our intuition here to determine $a_r^a(t)$ is to utilize martingales.

Following Cairns et al. (1998) and Macdonald et al. (2018), we define $M_r^{ad}(t)$ to be the counting process martingale of individual r in the alive-dead model as follows:

$$M_r^{ad}(t) = N_r^{ad}(t) - \int_0^t I_r^a(w)\mu_{r,w}^{ad}dw. \tag{6.10}$$

Then, since $e^{-\delta t}A_r^{ad}(t)$ is a deterministic (therefore previsible) function of calendar time t,

$$\int_0^\infty e^{-\delta t}A_r^{ad}(t)dM_r^{ad}(t) \tag{6.11}$$

is also a martingale. Defining $\mathcal{F}_0$ to be the complete life history of individual r at calendar time 0, we have:

$$E\left[\int_0^\infty e^{-\delta t}A_r^{ad}(t)dM_r^{ad}(t)\middle|\mathcal{F}_0\right] = 0 \tag{6.12}$$

which satisfies the equivalence principle. Note that what we do is much stronger than applying the equivalence principle to an individual in state a at calendar time zero. It ensures that $E[dL_r(t)] = 0$ for all t. This result enables us to state the premiums as follows:

$$a_r^a(t) = A_r^{ad}(t)\mu_{r,t}^{ad}, \tag{6.13}$$

which does not depend on the age (or the calendar time) at which insurance was purchased.

6.5 Insurance Losses in Multiple-State Models

Now we extend the alive-dead model (Figure 6.1) into a multiple-state model. Thus define a multiple-state model consisting of model states, which may include insured state(s), labelled by $0, 1, ..., \Psi$. Suppose individual r starts at state 0 at calendar time zero and might transfer to any model state, including an 'insured' state, after calendar time zero.

Let $N_r^{jk}(t)$ be the number of transitions of individual r from state j to state k for all $j{\neq}k$ up to and including calendar time t. Denote by $A_r^{jk}(t)$ (assumed to be

previsible like $A_r^{ad}(t)$ in the two-state model) the lump sum payment (possibly zero) payable to individual r under the transition $j \to k$ at calendar time t. $I_r^j(t)$ is defined to be an indicator function as follows:

$$I_r^j(t) = \begin{cases} 1, \text{ individual } r \text{ is in state } j \text{ at calendar time } t^-; \\ 0, \text{ otherwise.} \end{cases} \tag{6.14}$$

Define $a_r^j(t)$ (assumed to be previsible like $a_r^{ad}(t)$ in the two-state model) to be the rate per annum of premium payment (possibly zero) made if individual r is in state j at calendar time t^-.

The insurance loss arising from individual r in state j between calendar time t and $t + dt$ is then:

$$dL_r^j(t) = \left(\sum_{k:j \neq k} A_r^{jk}(t) dN_r^{jk}(t) - a_r^j(t) I_r^j(t) dt \right). \tag{6.15}$$

We showed, in equation (6.13), a premium rate function in the alive-dead model, which is motivated and obtained by utilizing martingales. Here, we follow the same motivation to determine $a_r^j(t)$, which a premium rate satisfying the actuarial equivalence principle (see Section 6.4.4). Define $M_r^{jk}(t)$ be the counting process martingale of individual r in state j as follows:

$$M_r^{jk}(t) = N_r^{jk}(t) - \int_0^t I_r^j(w) \mu_{r,w}^{jk} dw. \tag{6.16}$$

Then, in which we assume $e^{-\delta t} A_r^{jk}(t)$ is previsible,

$$\int_0^\infty e^{-\delta t} A_r^{jk}(t) dM_r^{jk}(t) \tag{6.17}$$

is a martingale; similarly, $\sum_{k:j \neq k} M_r^{jk}(t)$ is also a martingale at which we can conclude that:

$$E\left[\int_0^\infty \sum_{k;j \neq k} e^{-\delta t} A_r^{jk}(t) dM_r^{jk}(t) \middle| \mathcal{F}_0 \right] = 0 \tag{6.18}$$

and the rate of premium is then:

$$a_r^j(t) = \sum_{k:j\neq k} A_r^{jk}(t)\mu_{r,t}^{jk}, \tag{6.19}$$

that is, the sum of the intensities out of state j, weighted by the relevant benefits payable, which does not depend on the calendar time of entry to state j. Substituting this into equation (6.15) we see that the actuarial equivalence principle is satisfied as in Section 6.4.4.

6.6 Policy Values

A policy value for the insurance product is the expected present value (EPV) of future losses at calendar time t conditional on $\mathcal{F}_t$, the complete life history up to and including calendar time t. (Note that a policy value in life insurance mathematics logically is a measure of the expected present value of future cash flows at any calendar time conditioning on current calendar time. This measure enables the insurers to quantify the amount of reserve they should keep to meet their future liabilities to the policyholders.)

(a) Denote by V_t the policy value at calendar time t in the alive-dead model (given that being in state a at calendar time t):

$$V_t = E[L_r(t)|\mathcal{F}_t] = E\left[e^{\delta t}\int_t^\infty e^{-\delta w}dL_r(w)\middle| I_r^a(t) = 1\right]. \tag{6.20}$$

(b) Denote by V_t^j the policy value at calendar time t in state j in the multiple-state model (given that being in state j at calendar time t):

$$V_t^j = E[L_r^j(t)|\mathcal{F}_t] = E\left[e^{\delta t}\int_t^\infty \sum_l e^{-\delta w}dL_r^l(w)\middle| I_r^j(t) = 1\right]. \tag{6.21}$$

Policy values can also be computed as the solution of Thiele's differential equation. Norberg (1992) states Thiele's differential equation is a special case of equations (6.20) and (6.21) when the life history of a single individual is modelled in a

Markovian setting. In other words, as Cairns et al. (1998) express, conditioning on $\mathcal{F}_t$ is basically meant to be conditioning only on the occupied state at calendar time t. Let n be the term of the insurance policy.

(a) The policy value V_t which is conditional on being in state a at calendar time t in the alive-dead model, is given by the solution of Thiele's differential equation as follows:

$$\frac{d}{dt}V_t = \delta V_t + a_r^a(t) - \mu_{r,t}^{ad}(A_r^{ad}(t) - V_t). \tag{6.22}$$

(b) The policy value V_t^j which is conditional on being in state j at calendar time t in multiple-state models is given by the solution of the general form of Thiele's differential equations as follows:

$$\frac{d}{dt}V_t^j = \delta V_t^j + a_r^j(t) - \sum_{k:j\neq k} \mu_{r,t}^{jk}(A_r^{jk}(t) + V_t^k - V_t^j). \tag{6.23}$$

We use the fourth-order Runge-Kutta method (Section 3.5) to solve the general form of Thiele's equations (which can easily be applied to the solution of Thiele's differential equation of the alive-dead model). Note that we solve the equations backwards since the known boundary condition for the policy value, where the policy term is n years, is $V_{t+n}^j = 0$; thus $dt < 0$. Now, we can find the policy values at any calendar time t in any state j.

6.7 A Measure of Insurance Losses under Adverse Selection

Until now, we assumed that insurers would charge individuals some 'correct' premiums satisfying the equivalence principle, see equation (6.9). This assumes that the transition intensities used by the insurer in equation (6.19) are 'correct', in the sense of estimating the true biological and other risks represented by the intensities in the model.

Now suppose that instead of the 'correct' premium rate $a_r^j(t)$ associated with the true nature of individual r, the insurer charges a premium rate $\tilde{a}_r^j(t)$. In this case,

the insurer either will observe, if $\tilde{a}_r^j(t) > a_r^j(t)$, a profit:

$$E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - \tilde{a}_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right] < 0, \tag{6.24}$$

or, if $\tilde{a}_r^j(t) < a_r^j(t)$, a loss:

$$E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - \tilde{a}_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right] > 0. \tag{6.25}$$

In our model, adverse selection arises when the policyholder has information indicating elevated risk, which is not available to the insurer, such that $\tilde{a}_r^j(t) < a_r^j(t)$. This could occur in two ways:

(a) insurers are not allowed to access genetic test results;

(b) insurers are not allowed to access genetic test results and family history.

In both cases, equation (6.25) is satisfied. This leads us to define a measure of the expected individual cost of adverse selection over all states j (note that we assume everyone starts in state 0 at calendar time zero) as follows:

$$\frac{\begin{array}{l}E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - \tilde{a}_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right] \\ \quad - E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - a_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right]\end{array}}{E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\tilde{a}_r^j(t)I_r^j(t)dt\middle|\mathcal{F}_0\right]} \tag{6.26}$$

and since

$$E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - a_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right] = 0, \tag{6.27}$$

this measure is:

$$\frac{E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\left(A_r^{jk}(t)N_r^{jk}(t) - \tilde{a}_r^j(t)I_r^j(t)dt\right)\middle|\mathcal{F}_0\right]}{E\left[\int_0^\infty \sum_j \sum_{k:j\neq k} e^{-\delta t}\tilde{a}_r^j(t)I_r^j(t)dt\middle|\mathcal{F}_0\right]}. \tag{6.28}$$

This measure is interpreted as the uniform proportion by which premium rates would

have to increase to compensate the insurer for the losses due to adverse selection.

6.8 Multiple-State Multiple-Population Models

We so far presented 'a two-state model' in Section 6.4 and a 'multiple-state model' in Section 6.5 for the life history of a single individual occupying only one distinct population.

However, in our model, each individual will occupy one of the many populations (or sub-populations) that make up the whole population, in which each sub-population will have the same multiple-state space. And, the sub-populations of individuals will not be generally known to insurers. Our actuarial mathematics so far would be enough if the insurers could learn about the sub-populations of individuals. This material in this chapter needs to be extended to a 'multiple-state multiple-population model'. We do this in Appendix A because the rather elaborate notations of that model will be overcomplicated at this stage. Instead, in Chapter 7, we first present an example of 'multiple-state multiple-population models', which is the adverse selection model of HCM for life insurance. We then present the theory behind 'multiple-state multiple-population models' in Appendix A.

Chapter 7

An Adverse Selection Model of Hypertrophic Cardiomyopathy (HCM) for Life Insurance I: Model Specification

7.1 Introduction

In this chapter, we introduce a model for the adverse selection risk presented by HCM in a life insurance market and define the methodology for calculating insurance losses in respect of a single individual. (See also the necessary life insurance mathematics in Chapter 6 and Appendix A, which describe the technicalities of life insurance mathematics in this chapter; the computation and measure of the adverse selection costs in the whole population in Chapter 8; and, the results associated with the measure in Chapter 9).

In Section 7.2, we describe the adverse selection model as an extension of the testing model (Chapter 5). In Section 7.3, we give a brief introduction to our approach to calculating expected insurance losses in the adverse selection model. Sections 7.4 and 7.5 explain how insurers and individuals behave in the adverse selection model. Section 7.6 shows examples of life histories of HCM families in the adverse selection model. Section 7.7 presents the methodology to calculate insurance

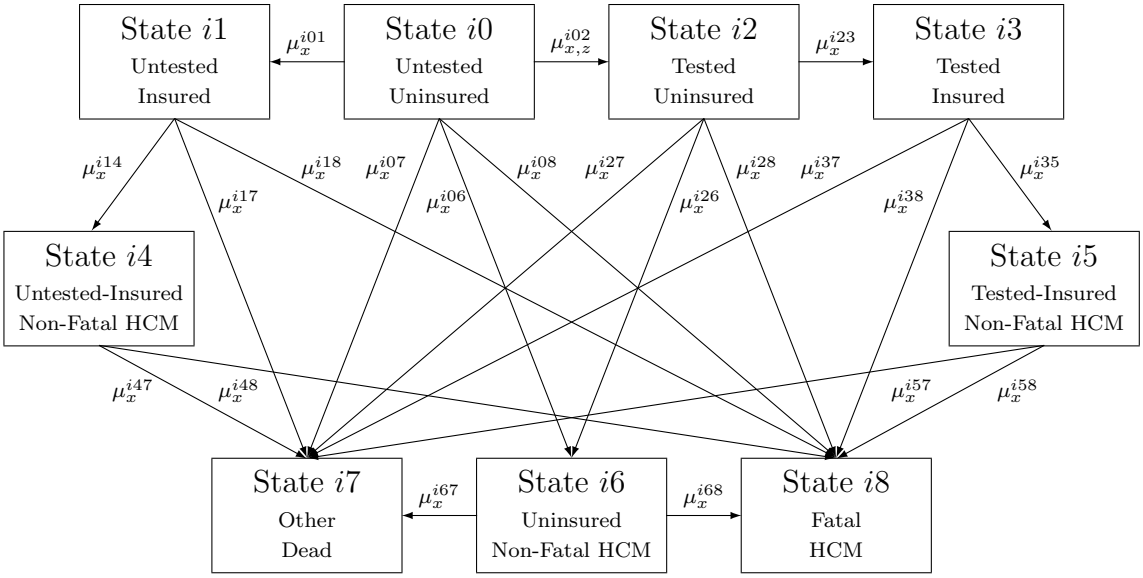


Figure 7.1: A mathematical model of adverse selection in HCM for a person in the ith of several sub-populations defined by HCM genotype in a life insurance market. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family.

losses in respect of a single individual in the adverse selection model. We discuss the adverse selection model in Section 7.8.

7.2 The Adverse Selection Model

In Chapter 5, we modelled the testing behaviour of individuals in HCM families. Our computational approach was to implement a simulation model. A key piece of family history information in the testing model was whether or not a proband exists in the family. Moreover, transition intensities from 'untested' states to 'tested' states of such a model also depended on duration since (if) a proband exists in the family. Therefore, the model was not Markov, even for an individual family member.

The adverse selection model extends the testing model by adding transitions into new 'insured' states, to represent the purchase of life insurance. It is shown in Figures 7.1, 7.2, and 7.3.

(a) Figure 7.1 shows a life history of an individual in a given sub-population in the adverse selection model, which has a bigger state space, compared to the testing model (Figure 5.1), due to adding insurance purchasing states.

(b) Figure 7.2 shows the nine sub-populations from the testing model, each expanded by the addition of the 'insured' states. The subdivision into the HCM population ($i = 2, 4, 6, 8$) and the non-HCM population ($i = 0, 1, 3, 5, 7$) is indicated by the dashed boxes. Figure 7.3 shows the same nine sub-populations, but subdivided into HCM families ($i = 1, 2, ..., 8$) and non-HCM families ($i = 0$). This makes the distinction between membership of the HCM population and membership of HCM families clear.

A family is represented by a collection of such models (Figures 7.1, 7.2, and 7.3), one per family member, which are linked by a common calendar time scale and genotypes inherited according to Mendel's law. The new states add two capabilities to the model.

- A person who is uninsured, and has not suffered death or an HCM-related event, can purchase insurance by moving into an 'insured' state. This can happen before genetic testing (state $i1$ in Figure 7.1) or after genetic testing (state $i3$ in Figure 7.1).

- Life insurance remains in force if an insured person suffers a non-fatal HCM-related event. Therefore, two additional states ($i4$ and $i5$ in Figure 7.1) are needed to allow for the insurance states if an HCM-related event occurs.

Note that there are no transitions from non-fatal HCM states into insured states because we assume that such a transition would be medically underwritten and would not lead to adverse selection.

We assume that the purchasing behaviour of carrier and non-carrier members in the 'untested uninsured' state of an HCM family can change as their testing behaviour changes. This means that we assume that insurance purchase rates before genetic testing depend on both the currently occupied state and family history (either there is a proband or there is no proband in the family) similar to the uptake rates of genetic testing. If individuals in HCM families take up genetic testing before purchasing insurance (in which the uptake of genetic testing depends on both the currently occupied state and family history), then, purchase rates depend on the

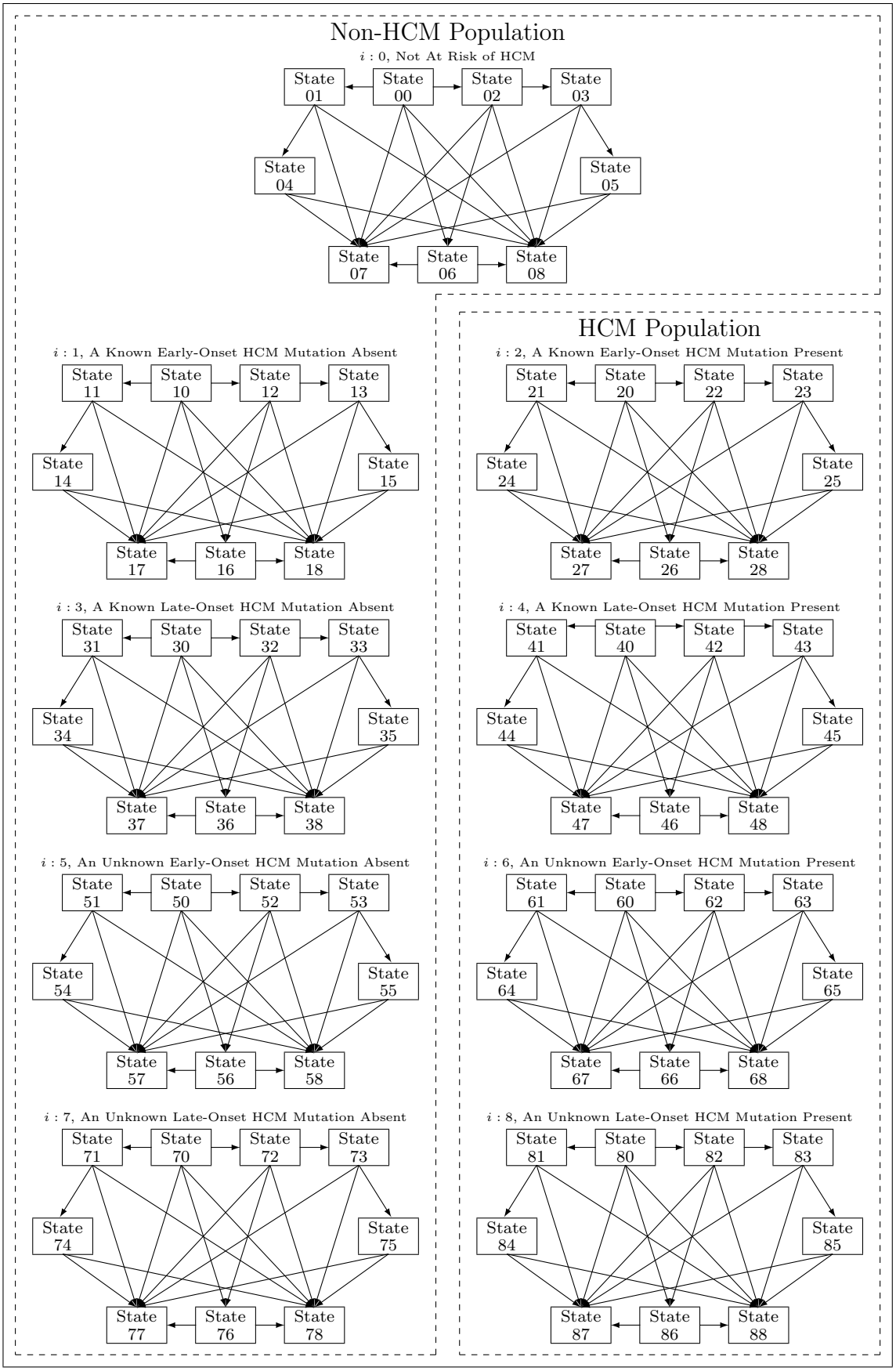


Figure 7.2: A mathematical model of adverse selection in HCM for a population with nine sub-populations associated with HCM genotype, in which each sub-population is classified as being a part of the HCM or non-HCM population, in a life insurance market.

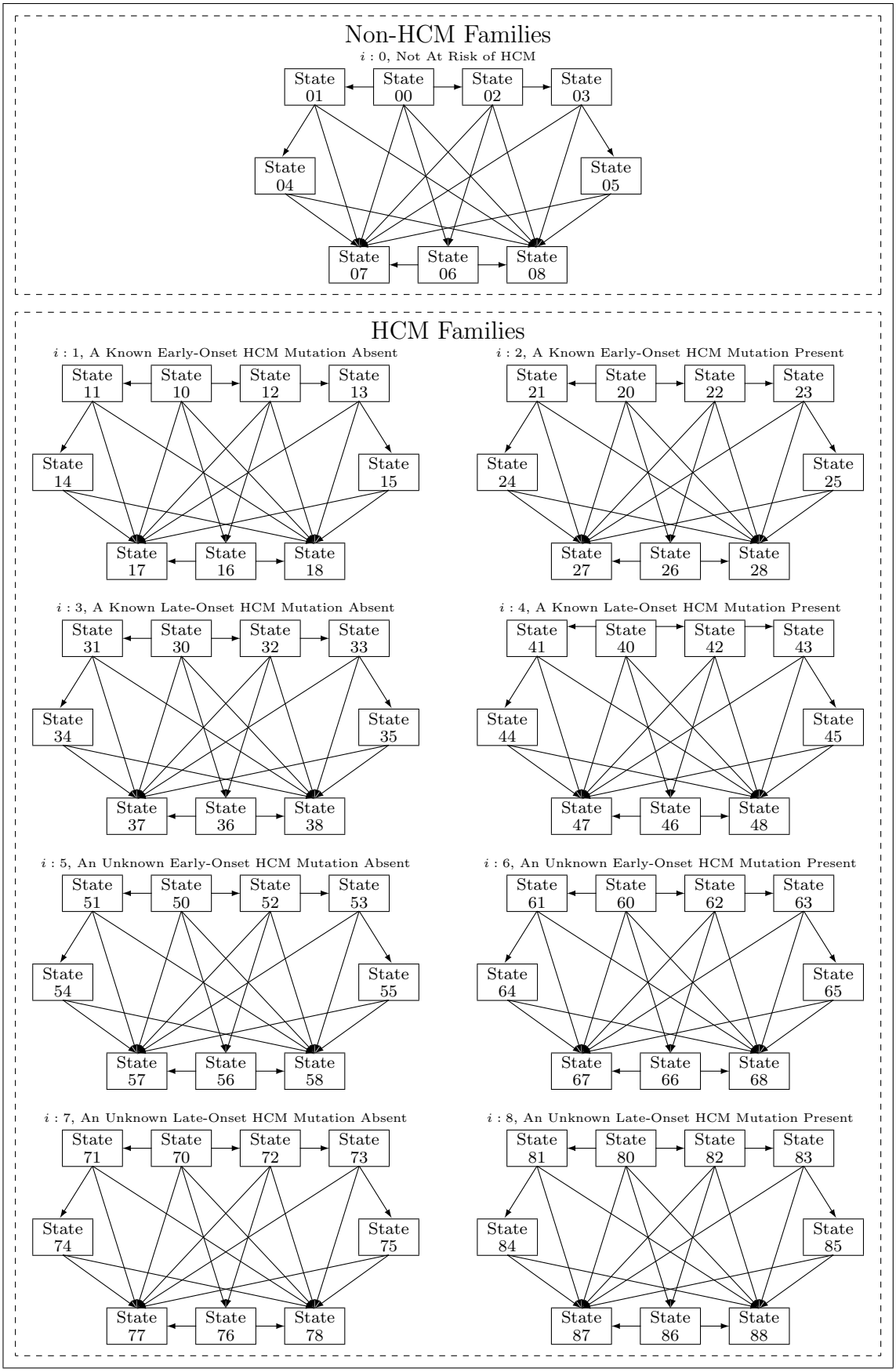


Figure 7.3: A mathematical model of adverse selection in HCM for a population with nine sub-populations associated with HCM genotype, in which each sub-population is classified as being a part of HCM or non-HCM families, in a life insurance market.

genetic test results, or the currently occupied state. The fundamental feature of the adverse selection model is the following:

(I) In the absence of any family history of HCM, individuals purchase insurance at some 'normal' rate. By default this rate always applies to the individuals in non-HCM families (in sub-population $i = 0$). Individuals purchase insurance between ages 20 and 60.

(II) In the presence of a family history of HCM, individuals purchase insurance based on the information available to them, namely the knowledge of their Mendelian risk if they have not been tested, or the genetic test result if they have been tested.

(III) In the presence of a family history of HCM, insurers charge premiums based on the information they have, or are allowed to use.

Adverse selection may arise because of discrepancies between the information used by individuals in (II) and by insurers in (III).

Note that not every transition, illustrated in Figures 7.2 and 7.3 (note that both figures refer to the same model, see point (b) in this section), is feasible at all times in all sub-populations. For example, genetic testing never occurs in non-HCM families (in sub-population $i = 0$), or in the sub-population representing individuals in HCM families carrying an unidentified, or 'unknown', mutation, $i = 5, 6, 7, 8$. For simplicity, we base our model on copies of the model in Figure 7.1, and impossible transitions have intensity zero.

7.3 Modelling Insurance Losses in the Adverse Selection Model

Another difference between the testing model and the adverse selection model is the way that cash flows arise in the latter. In this case, the insurer would have cash flows arising in two ways (Section 6.2):

(a) Paying a benefit/sum assured as an outgo, if the insured dies while insured during the policy term.

(b) Receiving a series of regular premium payments (actually a continuous cash-flow in our model) as an income, after insurance purchase and until the event of death or the expiry of the policy.

In the adverse selection model, we have a multiple-state multiple-population model, see Appendix A, but cash flows arise as in the two-state model only if a person is in one of the 'insured' states, see Section 6.4. In Sections 8.3.1 and 8.3.2, we implement this scheme to compute insurance losses in the adverse selection model.

7.3.1 Insurance Losses in HCM and Non-HCM Families

The testing model started with a fixed population at calendar time zero in the zeroth generation (see Section 5.3.4). Then, we paired off persons still alive at calendar time 20 in the zeroth generation and created HCM and non-HCM families. After that, we only focused on HCM families because uptake of genetic testing in non-HCM families is zero. However, we include all the individuals in non-HCM families in the adverse selection model since they will be 'normal' purchasers of insurance.

In Section 8.3.1, we obtain expected total insurance losses arising from non-HCM families by solving Thiele's differential equations (see Section 6.6 and Appendix A.1.2). Note that family history in non-HCM families has no effect on transition rates. Therefore, the life history of a single individual in non-HCM families is Markov with deterministic transition intensities and Thiele's differential equations are valid. The sum of the expected insurance losses in respect of each single individual will give us the expected total insurance losses in non-HCM families.

In Section 8.3.2, we obtain expected total insurance losses arising from HCM families. Insurance losses are much more complicated in HCM families partly because the life histories of single individuals are not Markov, and Thiele's equations are not valid. But the simulation model that we developed in Chapter 5 can be adapted to simulate cash flows in the adverse selection model. Insurance losses are

also complicated by the fact that individuals make insurance purchasing decisions and insurers set insurance premiums based on the information available to them.

As a result, we separately estimate expected total insurance losses in non-HCM families by solving Thiele's equations, and in HCM families by Monte Carlo simulation (Chapter 8), then we combine them.

7.3.2 How Do Adverse Selection Costs Arise?

Adverse selection gives rise to losses because the insurer calculates premium rates assuming that no adverse selection occurs. That is, that all insurance purchasing is 'normal' whereas in reality some insurance purchasing exploits information not available to the insurer. Measuring the resulting loss therefore involves two stages (see Section 6.7 and Appendix A.1.3):

(a) First, calculate the premium rates the insurer will charge based on available information and assuming no adverse selection. In our model, 'no adverse selection' equates to the absence of genetic testing, the non-Markov properties of the testing model described in Chapter 5 are absent, and premium rates can be calculated by solving the Kolmogorov equations under 'normal' purchasing behaviour (see Section 7.4.5).

(b) Second, calculate the expected present value (EPV) of the insurance losses that arise due to purchasers acting upon information not available to the insurer, while being charged the rates of premium calculated in (a) above.

7.4 Information and Decisions—Insurers

7.4.1 Underwriting Classes

The decisions made by insurers is based on the information available to them. In the basic scenarios this is the presence, or not, of a proband in the family. In other scenarios even this might not be known. Mathematically, we represent this quite simply by the definition of an underwriting class. At every point of time the insurer

allocates each individual to an underwriting class and the pricing decision is based on that. In our study, by default genetic test results are not disclosed to the insurer, but we assume family history might or might not be disclosed to the insurer. Note that, as a baseline assumption, 'negative' genetic test results are disclosed in our study based on the practice in the UK case, see Section 1.1. Then, the insurer will determine either one or two underwriting classes for the purchasers of insurance:

(a) In the case that insurers are allowed to use family medical history in underwriting, there will be two underwriting classes: persons with a family history of HCM (a proband in the family) and persons with no family history of HCM.

(b) In the case that insurers are not allowed to use family medical history in underwriting, there will be just one underwriting class, containing everyone. Note that the individuals tested negative will also be belong to this underwriting class.

The insurer will calculate premium rates based on that. This allows us to write down mathematical statements such as equation (7.1). See the mathematical derivation of equation (7.1) in Appendix A.1.1.

7.4.2 Age-Dependent Premium Rates

A common way to calculate premium rates in private insurance is as a level regular premium representing the fair price for the mortality risk that each individual brings to the insurance pool at the age when they purchase insurance. For this reason, insurers would charge different level premium rates at different ages of insurance purchase.

However, in our study, we do not use that method to determine premium rates because it would make the numerical solution of the problem messy, see Macdonald & Yu (2011). The reason is that our model includes the purchase of insurance as an event that may happen at any age between 20 and 60 (or not at all). If the insurer charged level premiums depending on the age at purchase, then at any given age x, persons in an insured state would be paying different premiums, depending on when they purchased their policies.

It is not impossible to use such a model, but it is very much simpler if all persons in an insured state at any given age x are paying the same rate of premium. Therefore, we define an underwriting class to be a set of insured states, based on family history information available to the insurer, such that all persons in any such state at a given age x are paying the same rate of premium. If individuals with different mortality rates are in the same underwriting class then the premium rate for that underwriting class must be a suitable some weighted average of the different rates of mortality. See Appendix A.1.1.

7.4.3 Calculating Premium Rates

We follow equation (A.6) to set up premium rates. This equation expresses the fact that if the insurer charges the insured a rate of premium equal to their mortality hazard rate multiplying the sum assured, the expected insurance loss between calendar time t and $t+dt$ would be zero, which will also satisfy the actuarial equivalence principle, that the expected present value of insurance losses at the inception of the policy should be zero.

Afterwards, we can charge premium rates based only on the mortality rates at age x (associated with purchased £1 sum assured), which are the weighted average transition intensities over all either the nine or the eight model sub-populations based on the information available to insurers (see Figure 7.4 and Section 7.4.4), from the 'insured model states' at age x (see the model states $i1$, $i3$, $i4$, and $i5$ in Figure 7.1) into the benefit claim model states, i.e 'dead states', in small time dt at age $x+dt$ (see the model states $i7$ and $i8$ in Figure 7.1).

Therefore, such a premium rate (for £1 sum assured) as a function of age x, denoted by $\phi_r^{C_r}(x)$, can be formulated as follows:

$$\phi_r^{C_r}(x) = \frac{\sum\limits_{ij\epsilon C_r} p_{i\ x} p_0^{i0j} (\mu_x^{ij7} + \mu_x^{ij8})}{\sum\limits_{ij\epsilon C_r} p_{i\ x} p_0^{i0j}}. \tag{7.1}$$

(See the proof of equation (7.1) in Appendix A.1.1.) Label r specifies an insurance buyer and all the quantities in the equation are associated with individual r. C_r

denotes the underwriting class to which insured individual r belongs. C_r consists of a set of pairs ij in which i is a sub-population and j is an insured sub-state in the ith sub-population. And, x is age in years; p_i is the prevalence rate at birth of being in the ith sub-population; ${}_xp_0^{i0j}$ is the (occupancy) probability of being in state ij at age x given in state $i0$ at age 0 (see Section 3.3 where we define occupancy probabilities in a multiple-state Markov model). Finally, μ_x^{ij7} or μ_x^{ij8} are the transition intensities from the insured states to the dead states ($ij \to i7$ or $ij \to i8$) in our model.

7.4.4 Premium Rates with Different Underwriting Classes

Different underwriting classes may arise from either having no proband or a proband in the family at the age of insurance purchase. Another difference will arise from gender because the cumulative probabilities of late-onset of HCM (see Figures 3.4 and 3.5) and the hazard rates of all-cause mortality are gender-related (see Figure 3.9). Note that we assume insurers are allowed to use gender to calculate premium rates (which has not been the case in the EU since 2012). Therefore, an underwriting class consists of a set of pairs ij for males and females separately. We determine the underwriting classes for male and female buyers separately under different scenarios as follows:

(a) An underwriting class is defined by the level of information that the insurer has, or is allowed to use. If that information, or prohibition on its use, means that the insurer cannot distinguish between presence in two or more insured states, these states must be in the same underwriting class.

(b) Assume that insurers are not allowed to use genetic test results (exceptionally negative test results are disclosed to the insurers in our model based on the UK case, see Sections 1.1 and 7.4.1) in underwriting. They may be allowed to use family history, which here means knowing whether or not there is a proband in the family. This creates two underwriting classes.

    (i) If there is no proband in the family when one of its members purchases insurance, the insurer knows only that the purchaser at the time of the

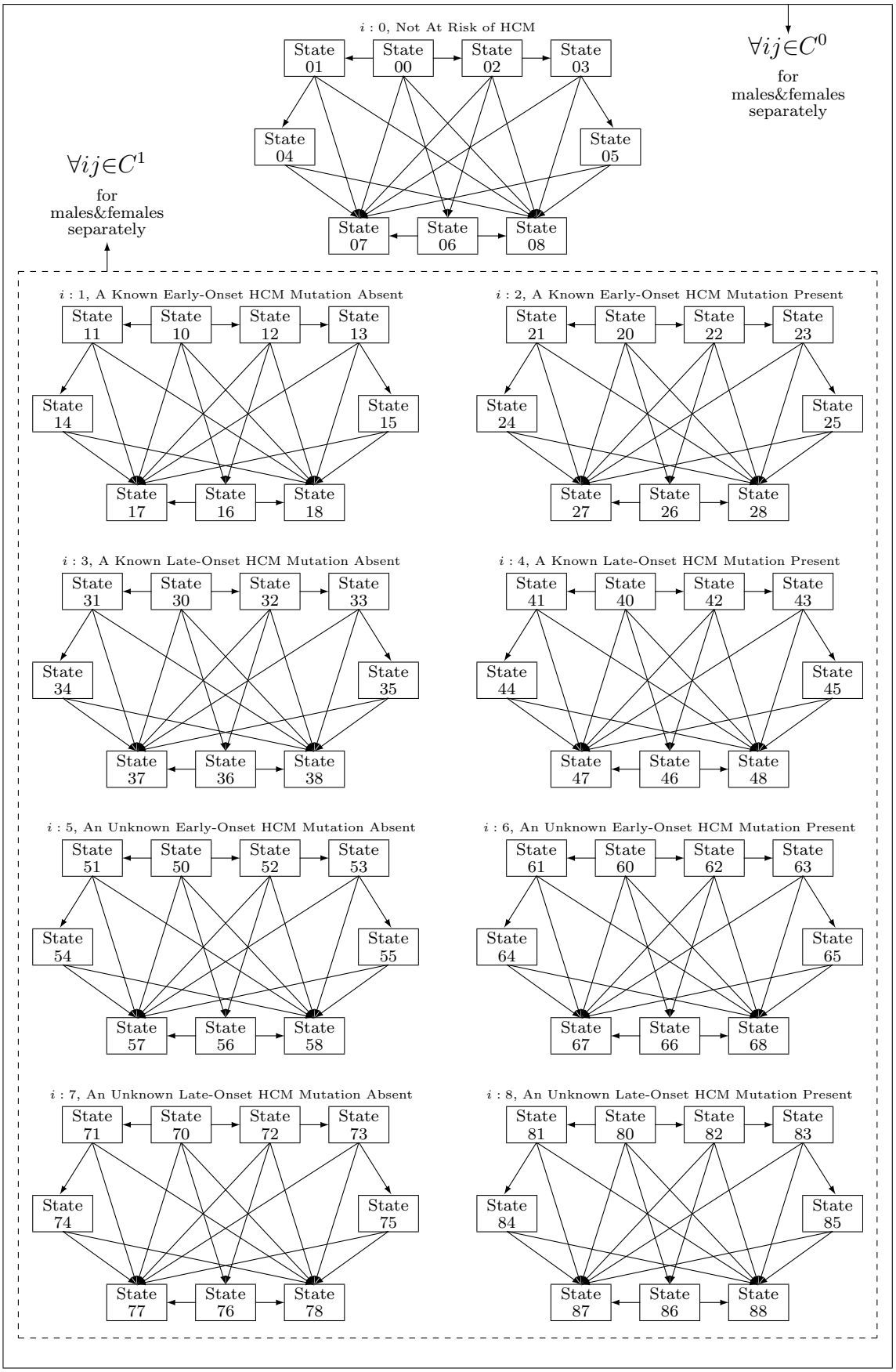


Figure 7.4: Involved model states to determine premium rates under different underwriting classes for males and females separately.

purchase is in one of the states $i1$ or $i3$ ($i = 0, 1, ..., 8$) but not which one. After the purchase of insurance, the insurer also does not know when (if) the individuals will transit into one of the 'insured non-fatal HCM' states $i4$ or $i5$. Therefore, the underwriting class contains all of these states. Denote this class by C^0 for males and females separately. See the solid box in Figure 7.4. Note that the persons tested negative (the UK case) are always in the underwriting class C^0 in our model.

(ii) If there is a proband in the family when one of its members purchases insurance, the insurer knows that the purchaser is in one of the states $i1$ or $i3$ ($i = 1, 2, ..., 8$) but is not in state 01 or 03. After the purchase of insurance, similar to point (i) above, individuals' transition into one of the 'insured non-fatal HCM' states $i4$ or $i5$ is unobserved by the insurer. Therefore, the underwriting class contains just these eight subpopulations. Denote this class by C^1 for males and females separately. See the dashed box in Figure 7.4.

If an individual labelled by r buys insurance at age x, the insurer determines the underwriting class to which they belong, denoted by C_r, for the purpose of deciding what premium rate in equation (7.1) to apply. If there is no proband in the family or individual r tested negative, $C_r = C^0$ (for males and females separately), otherwise, if there is a proband in the family, $C_r = C^1$ (for males and females separately). They are thereafter charged the rate of premium appropriate for that underwriting class regardless of subsequent events.

An individual is allocated to an underwriting class at the time of insurance purchase on the basis of family history at that time, see Figure 7.4. This is yet another reason why the adverse selection model is not Markov.

In earlier work (Gui et al. (2006) and Lu et al. (2007)) the development of a family history (e.g. two or more female first-degree relatives diagnosed with breast cancer before age 55) was represented by adding a 'family history' state to the model, transitions into this state being governed by a transition intensity obtained by building a separate external model to derive the distribution of the age at 'onset'

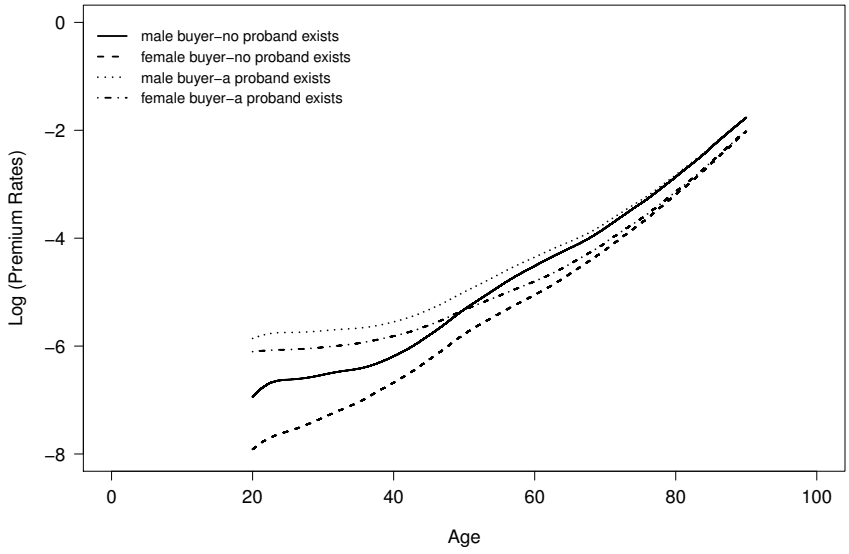


Figure 7.5: Estimated weighted average premium rates per unit benefit with the baseline assumptions (Table 7.1) under different underwriting classes. All rates converge to each other at older ages.

of a family history. Then, family history was embedded in the model in a way that preserved the Markov property. For our study of HCM, we needed to model families directly in order to model cascade genetic testing explicitly, so the earlier approach was not sufficient.

7.4.5 Calculated Premium Rates

We obtain the estimated premium rates shown in Figure 7.5 as follows:

(a) Determine the assumptions listed in Table 7.1 which extends the baseline assumptions in the epidemiological and testing model, see Tables 3.10 and 5.1, by adding the new assumptions associated with the insurance states in the adverse selection model: 0.05 normal purchase rate per annum (annual hazard rate of normal purchase of insurance) in a large market at ages 20–60 (Macdonald & Yu 2011); £1 normal sum assured; 0.05 force of interest per annum.

(b) Apply equation (7.1) with the assumptions represented in Table 7.1 under different underwriting classes.

Table 7.1: The assumptions underlying the premium rates shown in Figure 7.5.

Epidemiological Parameters	Table 3.10	Section 3.10
Prevalence of non-HCM mutations in the general population at age 20	0.998	Section 3.7.1
Prevalence of HCM mutations in the general population at age 20	0.002	Section 3.7.1
Prevalence of known early-onset mutations in the HCM population at birth	0.5	Section 3.7.2
Prevalence of known late-onset mutations in the HCM population at birth	0.1667	Section 3.7.2
Prevalence of unknown early-onset mutations in the HCM population at birth	0.25	Section 3.7.2
Prevalence of unknown late-onset mutations in the HCM population at birth	0.0833	Section 3.7.2
Penetrance of early-onset HCM at age 20	100%	Section 3.8.1
Penetrance of late-onset HCM at ages 20–70	Figure 3.4	Section 3.8.2
Hazard rate of fatal HCM per annum for all ages	0.0055	Section 3.9.7
Hazard rate of non-fatal HCM per annum for all ages	Table 3.6	Section 3.9.7
Hazard rate of all other death per annum for all ages	Figure 3.9	Section 3.10
Hazard rate of testing at all ages	0	Section 4.3.2
Hazard rate of normal purchase per annum at ages 20–60	0.05	Section 7.5.1
Benefit (Sum Assured)	£1	-
Force of interest per annum	0.05	-

(c) Calculate the occupancy probabilities to quantify premium rates as if genetic testing does not exist (see point (a) in Section 7.3.2). Therefore, they can be found by solving the Kolmogorov forward equations. See Sections 3.4 and 3.5 for the derivation and numerical solution of the Kolmogorov forward equations. As we noted in Sections 3.11 and 5.2.3, the Kolmogorov forward equations are numerically solved with time step 0.0005 years in this study.

We explain four components forming Figure 7.5 as follows:

- The solid line in the figure describes the premium rates at age x of a male buyer of insurance if there was no proband in the family at the time of the insurance purchase. This represents C^0 for males.
- The dashed line in the figure describes the premium rates at age x of a female buyer of insurance if there was no proband in the family at the time of the insurance purchase. This represents C^0 for females.
- The dotted line in the figure describes the premium rates at age x of a male buyer of insurance if there was a proband in the family at the time of the insurance purchase. This represents C^1 for males.
- The dotdash line in the figure describes the premium rates at age x of a female buyer of insurance if there was a proband in the family at the time of the insurance purchase. This represents C^1 for females.

Note that, following the UK case (Sections 1.1 and 7.4.1), we assume (as a baseline) the underwriting class of the individuals tested negative is C^0 (for both genders).

7.5 Information and Decisions—Individuals

7.5.1 'Information Classes' and Purchase Rates

The decisions made by individuals are based on the information available to them. We allocate each individual at any point in time to in one of four 'information classes' (as distinct from states in the multiple-state model). These are analogues of the insurers' underwriting classes.

(a) **Information Class $\boldsymbol{\zeta^n}$**. No information exists that suggests any HCM risk. This applies to non-HCM families always, and to members of HCM families before a proband has appeared. This also applies for the spouse of (when) the proband carrier parent, see (b) (i) below.

(b) **Information Class $\boldsymbol{\zeta^{50}}$**. An unaffected member of a family with a proband, who is at 50% risk of carrying a mutation. This applies to untested members of sub-populations $i = 1$ to $i = 8$ splitting into two cases.

(i) The proband is a parent. Then all untested children are in information class ζ^{50}, but the spouse of the proband is deemed to be in information class ζ^n.

(ii) The proband is a child. Then all untested children and both parents are in information class ζ^{50} if untested.

(c) **Information Class $\boldsymbol{\zeta^{100}}$**. A family member who was in information class ζ^{50} has been tested and carries a known mutation. This applies in sub-populations $i = 2$ and $i = 4$.

(d) **Information Class $\boldsymbol{\zeta^0}$**. A family member who was in information class ζ^{50} has been tested and does not carry a known mutation. This applies in sub-populations $i = 1$ and $i = 3$.

All of which moves us closer to a mathematical model of the costs of adverse selection. Our baseline assumptions are:

- If individuals in information class ζ^{100} buy insurance at more than the 'normal purchase rate', they will increase adverse selection costs.

- If individuals, who choose not to be tested, or are in sub-population $i = 5$ to $i = 8$, and are in information class ζ^{50} buy insurance more than at the 'normal purchase rate', those in sub-populations $i = 2$, $i = 4$, $i = 6$, or $i = 8$ will increase adverse selection costs while those in sub-populations $i = 1$, $i = 3$, $i = 5$, or $i = 7$ will decrease adverse selection costs.

Table 7.2: Baseline assumptions for insurance purchase (hazard) rates per annum depending on the information class in a large market. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation.

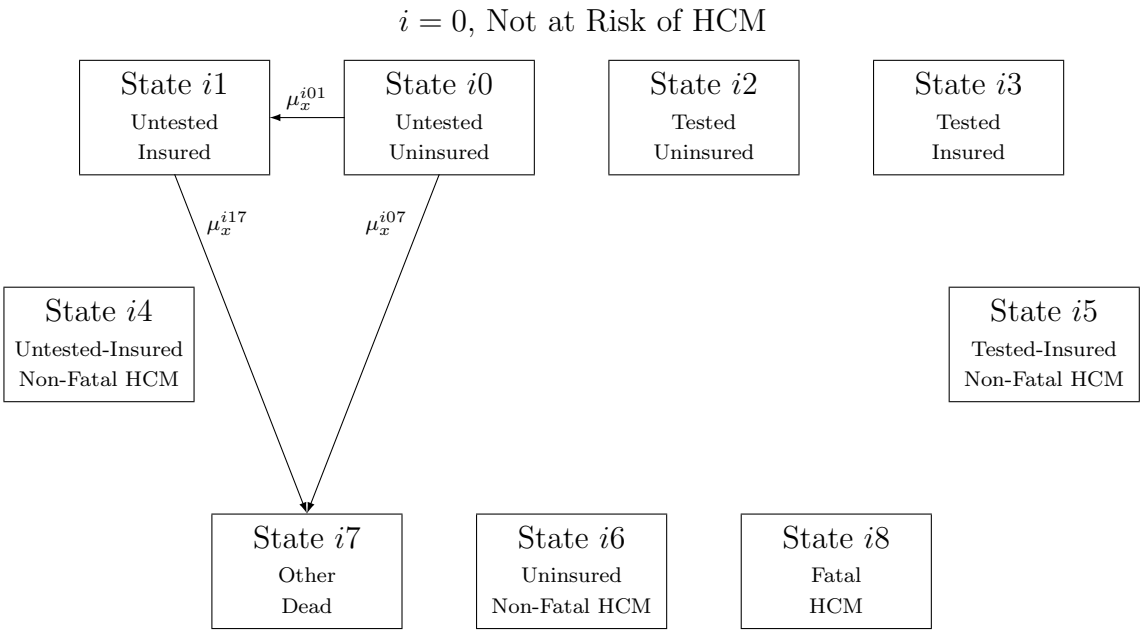
Adverse	Information Class			
Selection	ζ^n	ζ^{50}	ζ^{100}	ζ^0
None	0.05	0.05	0.05	0.05
Mild	0.05	0.1	0.1	0.05
Severe	0.05	0.25	0.25	0.05

- Note that adverse selectors (or individuals in class ζ^{100} and ζ^{50}) are assumed, as a baseline, to purchase at £1 normal sum assured. We consider, in Chapter 9, higher sums assured for the adverse selectors (either their purchase rate is normal or different than normal).

Table 7.2 shows the (baseline) annual purchase (hazard) rates based on the 'information classes' of individuals where 0.05 represents baseline normal purchase rate per annum (Table 7.1). These baseline purchase rates assume that individuals in class ζ^{50} behave (in terms of insurance purchasing) the same as class ζ^{100}. However, we do not really know whether or not they will behave in the same way as individuals in class ζ^{100} because they are not tested. In Chapter 9, we will relax this assumption by which individuals in class ζ^{50} might behave differently than class ζ^{100}.

7.5.2 Non-HCM Families: Risk Sub-population $i = 0$

Figure 7.6 shows the possible life histories of the individuals in non-HCM families (in sub-population $i = 0$). Note that these individuals never have genetic testing and they will always buy insurance at the 'normal purchase rate'. As a result, their life histories are Markov (see Sections 5.7 and 5.8). Then Figure 7.6 represents a multiple-state Markov model. So, occupancy probabilities in such a model can be found by solving the Kolmogorov forward equations (see Sections 3.4 and 3.5), and expected present value of insurance cash flows can be found by solving Thiele's equations (see Section 6.6 and Appendix A.1.2).



Fixed Rates:

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.
- $\boldsymbol{\mu_x^{ij7}}$ = annual mortality rate of all other causes at all ages.

Figure 7.6: A mathematical model of a life history of an individual who is a member of the $i = 0$ risk sub-population.

7.5.3 HCM Families: The Risk Sub-populations

Suppose HCM families contain a set of individuals $\mathcal{H} = \{1, 2, ..., \Omega\}$, partitioned into a collection of HCM families $F_1, ..., F_\omega$ where each $F_a \subset \mathcal{H}$. Let $m \in F_a$ be a member of an HCM family F_a.

7.5.3.1 Age of Individual m

Let b_m be the calendar time at the birth of individual m. The age of individual m at calendar time t, denoted by $x_m(t)$ is:

$$x_m(t) = t - b_m. \tag{7.2}$$

Calendar time is the natural timescale to use because parents and children have different ages at the same calendar time t. For simplicity, individuals in the same generation are assumed to be born at the same calendar time. See Section 5.3.1.

7.5.3.2 Generation of Individual m

Define:

$$G_m = \begin{cases} 1, \text{ if individual } m \text{ is a child, born at calendar time } b_m = 30, \\ 0, \text{ if individual } m \text{ is a parent, born at calendar time } b_m = 0. \end{cases} \tag{7.3}$$

Note that we assume that families have children (if their number is not zero) at calendar time 30. Therefore, $b_m = 30$ for children. See Section 5.3.1.

7.5.3.3 Gender of Individual m

Let g_m be the gender of individual m:

$$g_m = \begin{cases} 1, \text{ if individual } m \text{ is a female,} \\ 0, \text{ if individual } m \text{ is a male.} \end{cases} \tag{7.4}$$

We noted, in Section 5.3, that the gender of parents is deterministically determined at birth at calendar time $t = 0$, while we noted in Section 5.6.2 that the gender of each child in HCM families is randomly determined at birth at calendar time $t = 30$, with probability 0.5 of being male or female.

7.5.3.4 Sub-population of Individual m

Let i_m be the sub-population to which individual m belongs, where $i_m \in \{1, ..., 8\}$. Individual m is identified as a person in an HCM family, so $i_m \neq 0$. See Figure 7.3.

7.5.3.5 Information Classes and Decisions of Individual m

We can at any time assign to individual m who has not themselves suffered HCM, to one of the four information classes defined in Section 7.5.1; and, the testing and insurance purchasing decisions of individual m is based on that. We modelled the life history, including the testing behaviour, of an individual in an HCM family in Section 5.9.2. Here, additionally, we expand the life history to include insurance purchasing behaviour.

Table 7.3: Information classes for each family member in an HCM family in which one parent carries a known HCM mutation.

	Carrier Parent Proband		A Carrier Child Proband	
A Known Mutation	Before Testing	After Testing	Before Testing	After Testing
Non-carrier Parent	ζ^{n}	-	ζ^{50}	ζ^{0}
Carrier Parent	-	-	ζ^{50}	ζ^{100}
Non-carrier Child	ζ^{50}	ζ^{0}	-	-
Carrier Child	ζ^{50}	ζ^{100}	-	-
Non-carrier Sibling	-	-	ζ^{50}	ζ^{0}
Carrier Sibling	-	-	ζ^{50}	ζ^{100}

Table 7.4: Information classes for each family member in a family in which one parent carries an unknown HCM mutation.

An Unknown Mutation	Carrier Parent Proband	A Carrier Child Proband
Non-carrier Parent	ζ^{n}	ζ^{50}
Carrier Parent	-	ζ^{50}
Non-carrier Child	ζ^{50}	-
Carrier Child	ζ^{50}	-
Non-carrier Sibling	-	ζ^{50}
Carrier Sibling	-	ζ^{50}

(a) The information class of individual m, based on the events in the family, is shown in Tables 7.3 and 7.4. See also Table 7.2 for insurance purchase rates associated with each information class.

(b) The life history of individual m (based on the testing and purchasing decisions of individual m) is shown in Figures 7.7, 7.8, 7.9, and 7.10. Note that the testing of both parents is assumed to be a joint decision (see Section 5.9.2).

7.6 Simulated Life Histories in HCM Families

This section is similar to Section 5.9.3 in which we presented the simulated life histories of HCM families in the testing model. In this section, we demonstrate the simulated life histories corresponding to the adverse selection model, represented by the mean and standard deviation of 500 independent simulations of the occupancy probabilities in each state in the adverse selection model in respect of HCM families. They are obtained, represented in Figures 7.11 and 7.12, by the following assumptions for the adverse selection model parameters: the baseline assumptions in the testing model shown in Table 5.1; the average number of children per family (λ) assumed to be 1.8 (Section 5.6.1); all individuals (including adverse selectors) assumed to purchase insurance at the normal rate of 0.05 per annum (Section 7.5.1) (meaning that there is no adverse selection); and, time step 0.005 years used for the numerical computations in the simulation because it is found to be nearly as accurate as using 0.0005 years in Section 8.6.3 (Section 5.2.3).

In Figure 7.11, we present the mean of the occupancy probabilities of the simulated lives of the parents/the zeroth generation obtained from 500 independent simulations in each state in the adverse selection model in respect of the HCM families. In Figure 7.11:

- Differently from the testing model, the adverse selection model has insurance purchasing states.

- The conclusion is that the occupancy probabilities in state 3, tested insured (the middle left plot) are much less than that of in state 1, untested insured (the

upper middle plot). In other words, we do observe not so many individuals in the zeroth generation who take up genetic testing at ages 20–90 mainly because of the small number of child probands.

- No parent takes up genetic testing at ages 20–30 (state 2, tested uninsured state, the upper right plot) because we allow parents to have children at age 30. See Section 5.9.3.

In Figure 7.12, we present the mean of the occupancy probabilities of the simulated lives of the children/the first generation obtained from 500 independent simulations in each state in the adverse selection model in respect of the HCM families. In Figure 7.12:

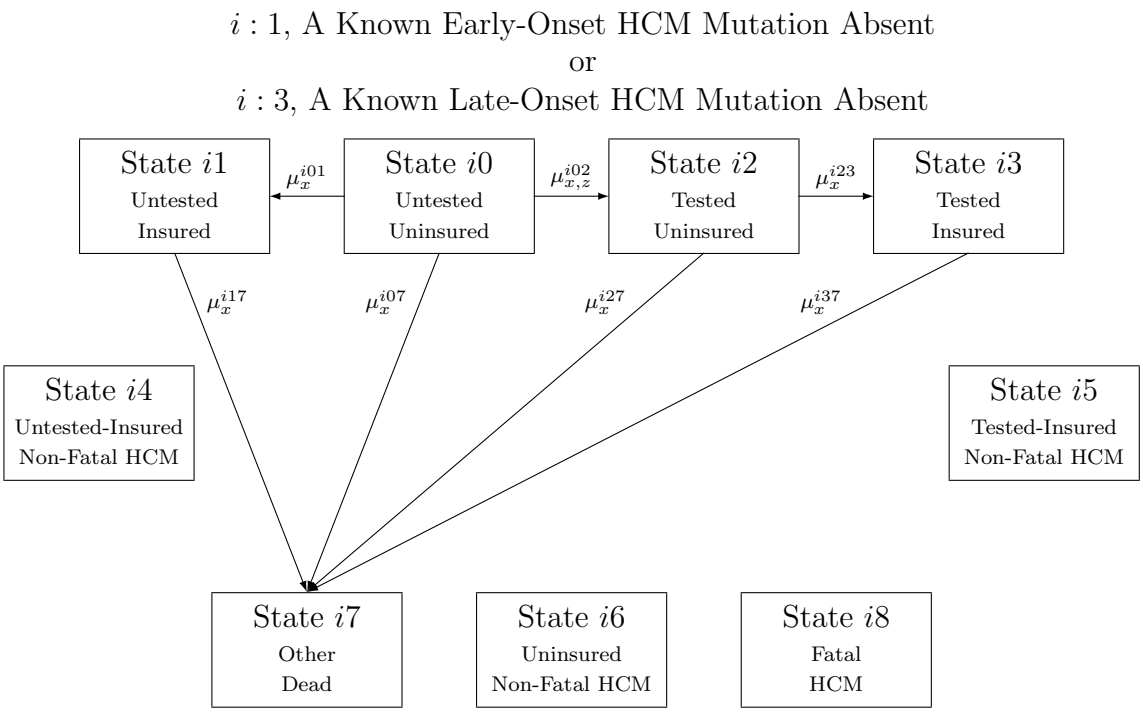
- The occupancy probabilities in insured states are zero up to age 20 because we assume the purchase rate is zero before age 20.

- The occupancy probabilities in state 3, tested insured (the middle left plot) are much higher compared to that of the parents' generation mainly because we observe more parent probands than child probands.

- Due to the infant mortality, there is a significant jump in state 0, untested uninsured (the upper left plot) correlated with state 7, other dead (the lower middle plot) just after age zero. See Section 5.9.3. Moreover, due to our conservative assumption that children whose carrier parents proband can be tested at age zero at calendar time 30, there is a significant jump in state 2, tested uninsured state (the upper right plot) at age zero. See Sections 4.3.2 and 5.9.3.

The standard deviations corresponding to Figures 7.11 and 7.12 are not shown since they are approximately zero at all ages.

Moreover, from the same reasoning in Sections 3.11 and 5.10, we do not make any sensitivity analysis for the simulated life histories in the adverse selection model since we aim to measure the insurance costs under adverse selection.



Fixed Rates:

- $\boldsymbol{\mu_x^{ij7}}$ = annual mortality rate of all other causes at all ages.

If no proband exists in family:

$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}} = 0$, uptake rate of testing per annum at any age.

If carrier parent becomes proband with a known mutation:

(a) *m is a spouse of carrier parent,* $\zeta_m = \zeta^n$*;*

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}} = 0$, uptake rate of testing per annum at any age.

(b) *m is a non-carrier child of carrier parent,* $\zeta_m = \zeta^{50}$*; or,* $\zeta_m = \zeta^0$*.*

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}}$ = normal uptake rate of testing per annum at all ages.
- $\boldsymbol{\mu_x^{i23}}$ = normal purchase rate (not change) per annum at ages 20–60.

If a carrier child becomes proband with a known mutation:

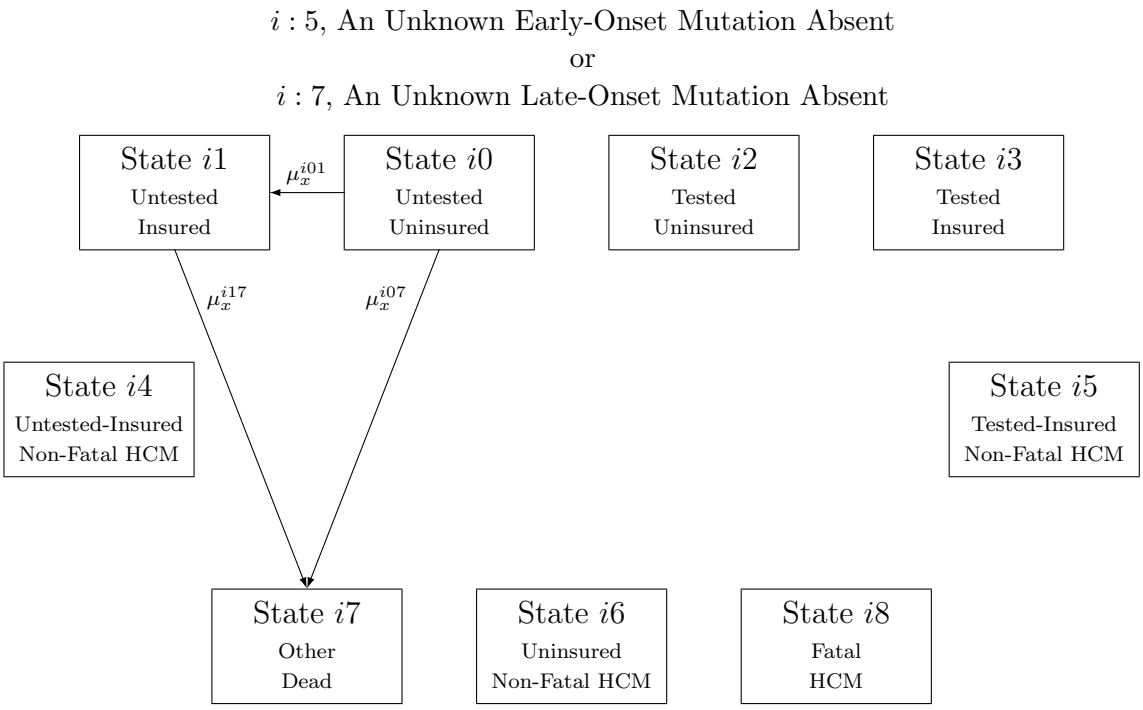
(a) *m is a spouse of carrier parent not tested nor become a subsequent proband; or, m is a non-carrier sibling of the carrier child,* $\zeta_m = \zeta^{50}$*; or,* $\zeta_m = \zeta^0$*.*

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}}$ = normal uptake rate of testing per annum at all ages.
- $\boldsymbol{\mu_x^{i23}}$ = normal purchase rate (not change) per annum at ages 20–60.

(b) *m is a spouse of carrier parent tested or become a subsequent proband,* $\zeta_m = \zeta^0$ *or* $\zeta_m = \zeta^n$*;*

- $\boldsymbol{\mu_x^{i01}} = \boldsymbol{\mu_x^{i23}}$ = normal purchase rate (not change) per annum at ages 20–60.

Figure 7.7: A mathematical model of a life history of an individual m, a non-carrier member in which one parent carries a known HCM mutation, in the $i = 1$ or $i = 3$ risk sub-populations in the adverse selection model of HCM for life insurance. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, and 7.4 for information classes.



Fixed Rates:

- $\boldsymbol{\mu_x^{ij7}}$ = annual mortality rate of all other causes at all ages.

If no proband exists in family:

$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.

If carrier parent becomes proband with an unknown mutation in family:

(a) *m is a spouse of carrier parent, $\zeta_m = \zeta^n$;*
- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.

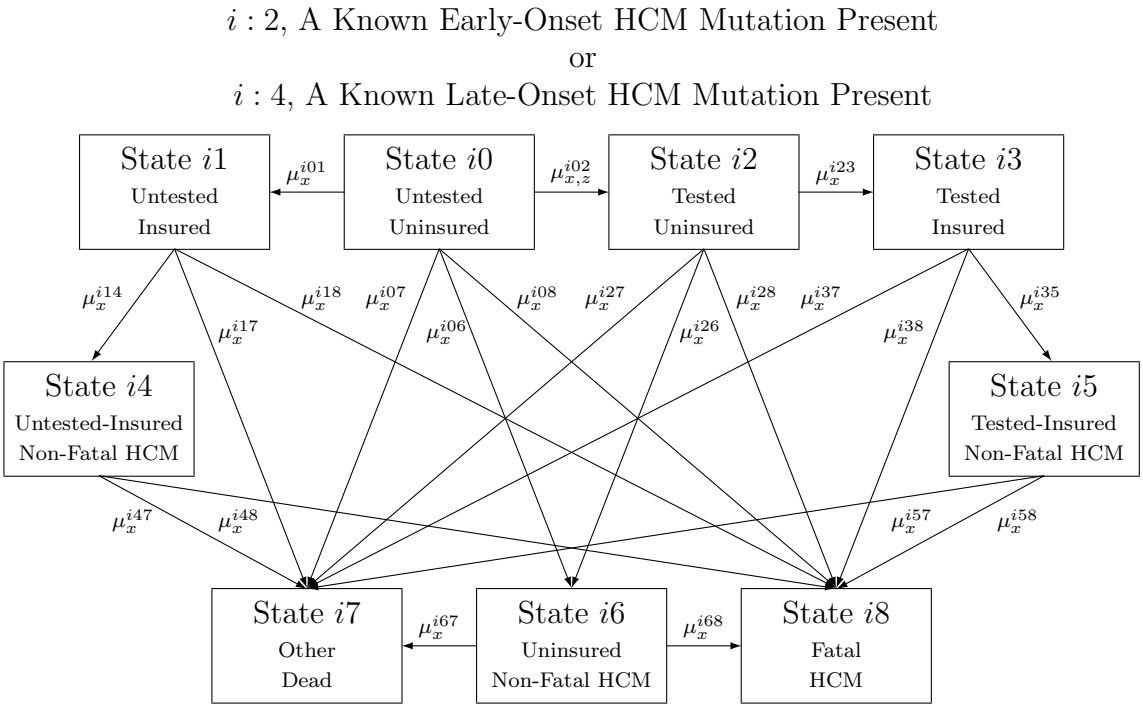
(b) *m is a non-carrier child of carrier parent, $\zeta_m = \zeta^{50}$;*
- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.

If a carrier child becomes proband with an unknown mutation in family:

(a) *m is a spouse of carrier parent not become a subsequent proband; or, m is a non-carrier sibling of the carrier child, $\zeta_m = \zeta^{50}$;*
- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.

(b) *m is a spouse of carrier parent become a subsequent proband, $\zeta_m = \zeta^n$;*
- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.

Figure 7.8: A mathematical model of a life history of an individual m, a non-carrier member in which one parent carries an unknown HCM mutation, in the $i = 5$ or $i = 7$ risk sub-populations in the adverse selection model of HCM for life insurance. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, and 7.4 for information classes.



Fixed Rates:

- $\boldsymbol{\mu_x^{i06} = \mu_x^{i14} = \mu_x^{i26} = \mu_x^{i35}}$ = proportioned, respective to penetrance of clinical HCM, non-fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{ij7}}$ = annual mortality rate of all other causes at all ages.
- $\boldsymbol{\mu_x^{i08} = \mu_x^{i18} = \mu_x^{i28} = \mu_x^{i38}}$ = proportioned, respective to penetrance of clinical HCM, fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{i48} = \mu_x^{i58} = \mu_x^{i68}}$ = fatal HCM rate per annum at all ages.

If no proband exists in family:

$\zeta_m = \zeta^n$;

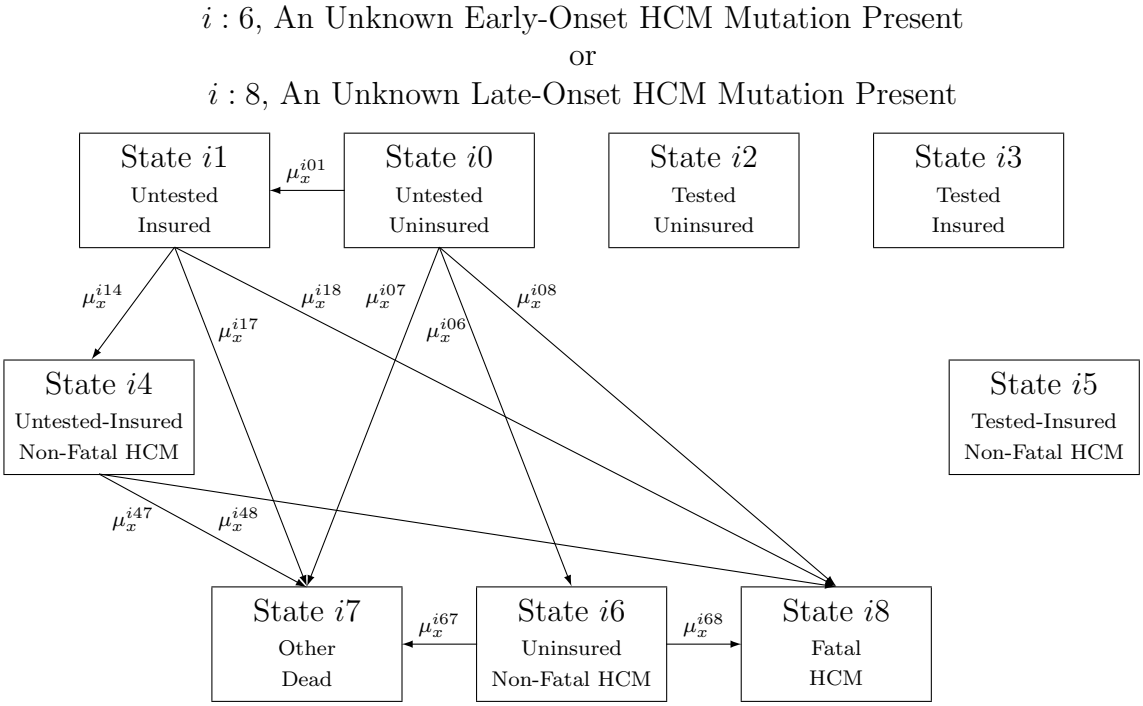
- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}} = 0$, uptake rate of testing per annum at any age.

If a proband exists (who not matter) with a known mutation in family:

$\zeta_m = \zeta^{50}$; or, $\zeta_m = \zeta^{100}$;

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.
- $\boldsymbol{\mu_{x,z}^{i02}}$ = normal uptake rate of testing per annum at all ages.
- $\boldsymbol{\mu_x^{i23}}$ = normal purchase rate (might change) per annum at ages 20–60.

Figure 7.9: A mathematical model of a life history of an individual m, a carrier member in which one parent carries a known HCM mutation, in the $i = 2$ or $i = 4$ risk sub-populations in the adverse selection model of HCM for life insurance. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, and 7.4 for information classes.



Fixed Rates:

- $\boldsymbol{\mu_x^{i06} = \mu_x^{i14}}$ = proportioned, respective to penetrance of clinical HCM, non-fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{ij7}}$ = annual mortality rate of all other causes at all ages.
- $\boldsymbol{\mu_x^{i08} = \mu_x^{i18}}$ = proportioned, respective to penetrance of clinical HCM, fatal HCM rate per annum at all ages.
- $\boldsymbol{\mu_x^{i48} = \mu_x^{i68}}$ = fatal HCM rate per annum at all ages.

If no proband exists in family:

$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (not change) per annum at ages 20–60.

If a proband exists (who not matter) with a known mutation in family:

$\zeta_m = \zeta^{50}$;

- $\boldsymbol{\mu_x^{i01}}$ = normal purchase rate (might change) per annum at ages 20–60.

Figure 7.10: A mathematical model of a life history of an individual m, a carrier member in which one parent carries an unknown HCM mutation, in the $i = 6$ or $i = 8$ risk sub-populations in the adverse selection model of HCM for life insurance. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, and 7.4 for information classes.

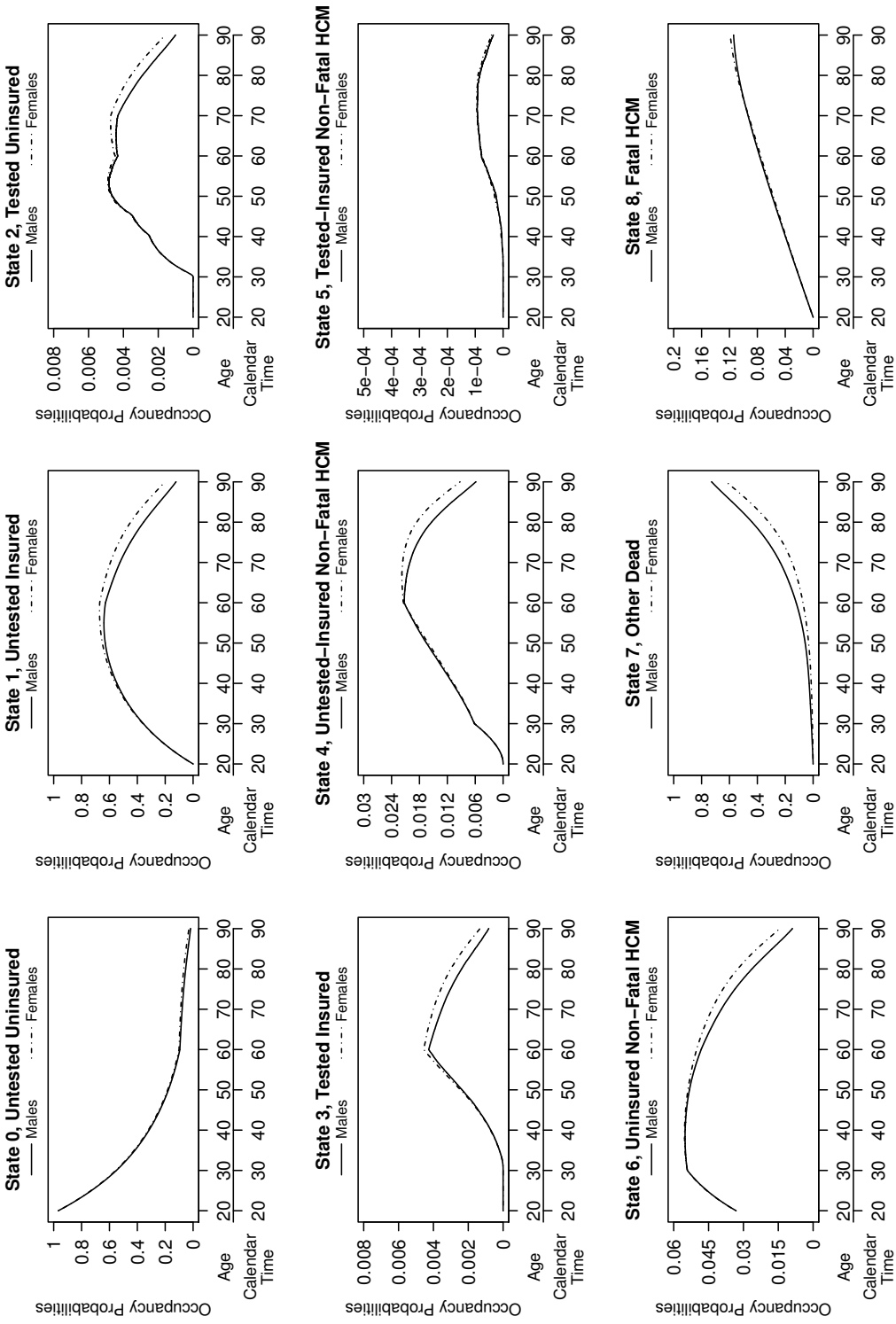


Figure 7.11: Mean of occupancy probabilities in each state in the adverse selection model under baseline assumptions obtained from 500 independent simulations in the zeroth generation in the HCM families. Note that the vertical axes have different scales.

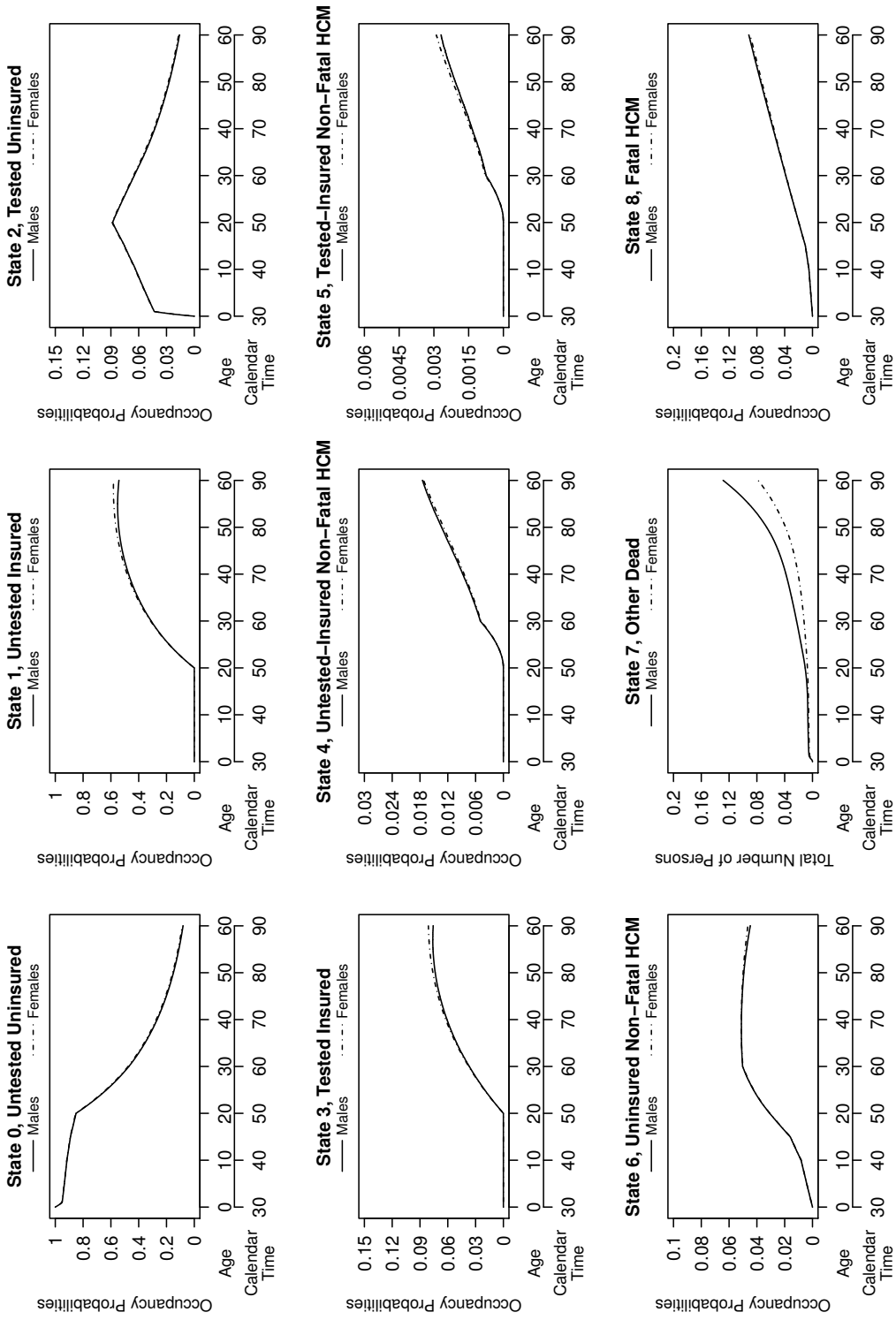


Figure 7.12: Mean of occupancy probabilities in each state in the adverse selection model under baseline assumptions obtained from 500 independent simulations in the first generation in the HCM families. Note that the vertical axes have different scales.

7.7 Individual Insurance Cashflows and Losses

7.7.1 Insurance Purchase by Individual m

If individual m buys insurance at random calendar time τ_m, then the age at the purchase time is:

$$x_m(\tau_m) = \tau_m - b_m. \tag{7.5}$$

However, if individual m never buys insurance, then define:

$$\tau_m = \infty. \tag{7.6}$$

All insurance contracts in our model are term life insurance expiring at age 60. We assume that all such that insurance policies are sold between ages 20 and 60, so if $\tau_m < \infty$ we have:

$$20 \leq x_m(\tau_m) \leq 60. \tag{7.7}$$

7.7.2 Underwriting Class of Individual m

The underwriting class of individual m is determined based on information available to the insurer at the calendar time of the insurance purchase, τ_m. A moratorium or family history may shape the information available in respect of individual m.

(a) If genetic test results are undisclosed to the insurer (exceptionally negative test results, see Sections 1.1 and 7.4.1), but family history is disclosed, the key information of interest to the insurer is whether or not a proband exists in the family of individual m at time τ_m. See the C^1, see the dashed box in Figure 7.4, premium rates for males and females separately.

(b) If both genetic test results and family history are undisclosed to the insurer, the insurer regards each individual as being in any of the sub-population with the probabilities used to calculate the C^0, see the solid box in Figure 7.4, premium rates for males and females separately.

Being a proband in our model is only possible with the transitions to non-fatal

HCM or fatal HCM model states (see the model states $i4$, $i5$, $i6$, or $i8$ in Figure 7.1). Therefore, we can observe if there is a proband in the family of individual m at calendar time t by counting the number of transitions into these states made by all the members of the family F_a including individual m as follows:

$$\Psi_m(t) = \sum_{r \in F_a} \sum_{j=0}^{8} \left(N_r^{i_r j4}(t) + N_r^{i_r j5}(t) + N_r^{i_r j6}(t) + N_r^{i_r j8}(t) \right). \tag{7.8}$$

If individual m has a proband in the family at (or rather, just before) the calendar time of the purchase of insurance, τ_m, a useful indicator function of that event is:

$$Y_m(\tau_m) = \begin{cases} 1, \ \Psi_m(\tau_m^-) > 0 \ \ \text{(a proband exists in family } F_a \text{ at time } \tau_m^-), \\ 0, \ \Psi_m(\tau_m^-) = 0 \ \ \text{(no proband exists in family } F_a \text{ at time } \tau_m^-). \end{cases} \tag{7.9}$$

Then, C_m (for males and females separately) is the underwriting class of individual m, already described in Section 7.4.4 as follows:

(a) If genetic test results are undisclosed to insurers, but family history is disclosed (or individual m tested negative) the underwriting class of uninsured individual m based on the information available to the insurer is:

$$C_m = \begin{cases} C^0, & \text{if } Y_m(\tau_m) = 0 \text{ or individual } m \text{ tested negative}, \\ C^1, & \text{if } Y_m(\tau_m) = 1. \end{cases} \tag{7.10}$$

(b) If both genetic test results and family history are undisclosed to insurers, the underwriting class of individual m is always $C_m = C^0$ (it does not matter which family member that individual m represents).

Note that C_m is determined at the time of the insurance purchase. It stays the same during the policy term. This definition assumes that negative tests are disclosed to the insurer and may be used based on the case under the UK moratorium. See Sections 1.1 and 7.4.1.

7.7.3 Premium Rates of Individual m

Each person's premium rates, see equation (7.1), are determined at the time of the purchase of insurance, τ_m. They are functions of age given by equation (7.1) and are denoted as follows:

$$\phi_m^{C_m}(x_m(t)), \quad \tau_m < t, \tag{7.11}$$

for males and females separately.

7.7.4 Premium Income from Individual m

See Section 6.4.2 for notation for this section. The individuals who moved into one of the insured states in our model (states $i1$, $i3$, $i4$, or $i5$ in Figure 7.1) pay the premiums determined in Section 7.7.3, as long as they are alive, or until their policy ends.

Define an indicator function as follows:

$$I_m^a(t) = \begin{cases} 1, \text{ the individual } m \text{ is insured and alive at calendar time } t^-. \\ 0, \text{ otherwise.} \end{cases} \tag{7.12}$$

Therefore $I_m^a(t)$ is equal to 1 if and only if the individual m pays premiums at calendar time t, continuously at annual rate $\phi_m^{C_m}(x_m(t))$ (for males and females separately), see equation (7.11). Then, the present value at age 20 of the premium payments of individual m is:

$$a_m = \int_0^\infty \phi_m^{C_m}(x_m(t)) e^{-\delta(x_m(t)-20)} I_m^a(t) dt \tag{7.13}$$

where δ is the constant rate of force of interest per annum.

7.7.5 Benefit Outgo in Respect of Individual m

See Section 6.4.1 for notation for this section. If an individual who is insured, moves into the one of the dead states during the policy duration (states $i7$ or $i8$ in Figure 7.1), the benefit will be paid immediately. We assume that the benefit amount is

$A_m(t)$.

Define $N_m^{i_m jk}(t)$ to be the total number of transitions made by individual m from state $i_m j$ to state $i_m k$ up to and including time t. Then the total number of insurance claims made by individual m up to and including time t, denoted by $N_m^{ad}(t)$, is.

$$N_m^{ad}(t) = \sum_{j=1,3,4,5} \left(N_m^{i_m j7}(t) + N_m^{i_m j8}(t) \right) \tag{7.14}$$

(Note that $N_m^{ad}(t)$ takes the values 0 or 1). Moreover, $A_m(t)N_m^{ad}(t)$ is the total amount of benefit paid to the insured individual m up to and including time t. Therefore, the present value at age 20 of the lump sum payment of $A_m(t)$ is:

$$\mathbb{A}_m = \int_0^\infty A_m(t) e^{-\delta(x_m(t)-20)} dN_m^{ad}(t). \tag{7.15}$$

7.7.6 The Insurance Loss in Respect of Individual m

See Section 6.4.3 for notation for this section. Then the present value at age 20 of the future losses from the insured individual m is:

$$L_m = \mathbb{A}_m - a_m. \tag{7.16}$$

If there is no adverse selection, the expected present value at age 20 of the future losses from the insured individual m is zero (from the actuarial principle of premium calculations):

$$E[L_m] = 0. \tag{7.17}$$

If there is adverse selection in our model, the expected present value at age 20 of the future losses from the insured individual m is not equal to zero (the actuarial principle of premium calculations does not hold):

$$E[L_m] \neq 0. \tag{7.18}$$

7.8 Discussion

This chapter modelled the life history of a single individual in respect of genetic testing and insurance purchasing behaviour. It also presented the computation of the insurance losses in respect of a single individual. We are ready to discuss, in Chapter 8, how we obtain the insurance losses from the whole population and measure them under adverse selection. Afterwards, we will be ready to see the estimated adverse selection costs in the whole population in Chapter 9.

Chapter 8

An Adverse Selection Model of Hypertrophic Cardiomyopathy (HCM) for Life Insurance II: Monte Carlo Simulation in HCM Families

8.1 Introduction

In Chapter 5, we modelled a general population composed of independent nuclear families and called the testing model. Each nuclear family contained two generations: parents as the zeroth generation and their children as the first generation, the latter possibly empty. Individuals but not families exist at calendar time zero (an equal number of males and females at birth). The males and females alive at calendar time 20 are paired off to create families. Two types of families arise: HCM and non-HCM families. See Figure 5.3. In Chapter 7, we extend the testing model with insurance purchasing states (called the adverse selection model) and discuss the computation of individual insurance losses from these families. (See the sub-populations in HCM and non-HCM families in the adverse selection model in Figure

7.3). We obtain separately expected total insurance losses from HCM and non-HCM families because we use different numerical techniques to measure expected total insurance losses from both types of families; afterwards, we aggregate them to find expected aggregated losses in the whole population. See Section 7.3.1.

In Section 8.2, we implement the overall methodology to compute the expected present value (EPV) of aggregated insurance losses in the whole population. In Sections 8.3 and 8.4, we explain the computation and combination of the EPVs of total losses from non-HCM and HCM families in detail. In Section 8.5, we give a list of baseline assumptions in the adverse selection model. The EPV of aggregated losses with the baseline assumptions in our model is demonstrated and discussed under no adverse selection in Section 8.6 and under some adverse selection scenarios in Section 8.7. In Section 8.8, we introduce a measure of the adverse selection costs and then present the corresponding results under many different adverse selection scenarios in Chapter 9.

8.2 Monte Carlo Simulation

1. Decompose the present value (PV) of aggregated losses at age 20 in the whole population denoted by L into two pieces:

$$L = L^{\mathcal{H}} + L^{\mathcal{N}}, \tag{8.1}$$

in which we define

   (a) $L^{\mathcal{H}}$, the PV of total losses at age 20 from HCM families.

   (b) $L^{\mathcal{N}}$, the PV of total losses at age 20 from non-HCM families.

2. Denote $E[L]$, the EPV of the aggregated losses at age 20 in the whole population, defined as follows:

$$E[L] = E\left[L^{\mathcal{H}} + L^{\mathcal{N}}\right] \tag{8.2}$$

in which

(a) $E[L^{\mathcal{H}}]$ is computed by Monte Carlo simulation of family life histories. See Section 8.3.2.

(b) $E[L^{\mathcal{N}}]$ is computed by solving Thiele's equations numerically. See Section 8.3.1.

8.3 Total Insurance Losses

8.3.1 Non-HCM Families

As we describe, in Sections 5.7, 5.8, and 7.5.2, the life history of a single individual in non-HCM families (or in risk sub-population $i = 0$) is always Markov. Therefore, we can calculate the expected present value (EPV) of future losses per an individual for males and females separately by solving Thiele's differential equations.

(a) If we set up the lump sum benefit £0 in Thiele's equations, then we obtain the EPV of future premium income in respect of a single individual for males and females separately.

(b) Similarly, if we set up the premium income to zero in Thiele's equations, we obtain the EPV of future benefit outgo in respect of a single individual for males and females separately.

See, in Appendix A.1.2, the general form of Thiele's equations for a multiple-state multiple-population Markov model. We denoted by V_t^{ij} the expected present value (the prospective policy value) of future losses in respect of an individual in state ij at calendar time t (given that being in state ij at calendar time t). We calculate the following in respect of an individual alive in state 00 at age 20 at calendar time t (given that being in state 00 at age 20 at calendar time t) (note that parents are age 20 at calendar time 20, but children are age 20 at calendar time 50, see Section 5.3):

- the EPV of total future losses from an individual (for males and females separately) at age 20, denoted by V_t^{00},

- the EPV of total future premium income from an individual (for males and females separately) at age 20, denoted by I_t^{00},

- the EPV of total future benefit outgo from an individual (for males and females separately) at age 20, denoted by O_t^{00}.

Then, we calculate the following in respect of non-HCM families consisting of $\mathcal{N}$ individuals alive at age 20:

- the EPV of total future losses at age 20 from non-HCM families, denoted by $E\left[L^{\mathcal{N}}\right]$, is:

$$E\left[L^{\mathcal{N}}\right] = \mathcal{N} \times V_t^{00}, \tag{8.3}$$

- the EPV of total future premium income at age 20 from non-HCM families, denoted by $E[I^{\mathcal{N}}]$, is:

$$E\left[I^{\mathcal{N}}\right] = \mathcal{N} \times I_t^{00}, \tag{8.4}$$

- the EPV of total future benefit outgo at age 20 from non-HCM families, denoted by $E[O^{\mathcal{N}}]$, is:

$$E\left[O^{\mathcal{N}}\right] = \mathcal{N} \times O_t^{00}. \tag{8.5}$$

The EPVs of total future losses, total future premium income, and total future benefit outgo at age 20 from non-HCM families above will be aggregated with the corresponding quantities from HCM families which are obtained in Section 8.3.2.

8.3.2 HCM Families

As we describe, in Section 5.6, HCM families have children (possibly zero) if both parents are alive at calendar time 30. An HCM family is a nuclear family: a mutation carrier parent, spouse of the carrier parent (assumed to be a non-carrier), and their children (if any). The total number of children in a family is assumed to be Poisson distributed with parameter λ. Also, each child inherits the mutation with probability 0.5 based on Mendel's law. In Sections 5.6, 5.9, and 7.5.3, we presented what are the characteristics of an HCM family and how we simulate the life histories of HCM

family members. Section 7.7 deals with insurance cashflows and losses arising from an individual m in an HCM family. Then, we calculate the expected present value of total future losses in HCM families which we compute by Monte Carlo simulation of family life histories (see Section 7.7 for the following notations) as follows:

(a) Let $L_{m,e}$ denote the present value of total future losses at age 20 in respect of an individual m in HCM families under the Monte Carlo simulation e where $e \in \{0, 1, ..., \mathcal{E}\}$, see equation (7.16). We estimate $E\left[L^{\mathcal{H}}\right]$ by the Monte-Carlo estimate, denoted by $\hat{E}\left[L^{\mathcal{H}}\right]$, as follows:

$$\hat{E}\left[L^{\mathcal{H}}\right] = \frac{1}{\mathcal{E}} \sum_{e=1}^{\mathcal{E}} \left(\sum_{m=1}^{\Omega_e} L_{m,e} \right). \tag{8.6}$$

Note that $\hat{E}\left[L^{\mathcal{H}}\right]$ converges to the (true) value of $E\left[L^{\mathcal{H}}\right]$ as $\mathcal{E} \to \infty$. Note also that the total number of individuals in HCM families depends on the simulation, so we denote it by Ω_e.

(b) Let $a_{m,e}$ denote the present value of total future premium income at age 20 in respect of an individual m in HCM families under the Monte Carlo simulation e where $e \in \{0, 1, ..., \mathcal{E}\}$, see equation (7.13). We estimate $E\left[I^{\mathcal{H}}\right]$ by the Monte-Carlo estimate, denoted by $\hat{E}\left[I^{\mathcal{H}}\right]$, as follows:

$$\hat{E}\left[I^{\mathcal{H}}\right] = \frac{1}{\mathcal{E}} \sum_{e=1}^{\mathcal{E}} \left(\sum_{m=1}^{\Omega_e} a_{m,e} \right), \tag{8.7}$$

Note that $\hat{E}\left[I^{\mathcal{H}}\right]$ converges to the (true) value of $E\left[I^{\mathcal{H}}\right]$ as $\mathcal{E} \to \infty$.

(c) Let $\mathbb{A}_{m,e}$ denote the present value of total future benefit outgo at age 20 in respect of an individual m in HCM families under the Monte Carlo simulation e where $e \in \{0, 1, ..., \mathcal{E}\}$, see equation (7.15). We estimate $E\left[O^{\mathcal{H}}\right]$ by the Monte-Carlo estimate, denoted by $\hat{E}\left[O^{\mathcal{H}}\right]$, as follows:

$$\hat{E}\left[O^{\mathcal{H}}\right] = \frac{1}{\mathcal{E}} \sum_{e=1}^{\mathcal{E}} \left(\sum_{m=1}^{\Omega_e} \mathbb{A}_{m,e} \right). \tag{8.8}$$

Note that $\hat{E}\left[O^{\mathcal{H}}\right]$ converges to the (true) value of $E\left[O^{\mathcal{H}}\right]$ as $\mathcal{E} \to \infty$.

8.4 Aggregated Insurance Losses

Here we explain how we combine the results from the HCM and non-HCM families:

(a) EPVs at age 20 in the non-HCM families are obtained by solving Thiele's equations and multiplying by the sum of:

(i) $2,462,620$ males and $2,470,647$ females surviving individuals at age 20 in the zeroth generation; and,

(ii) $2,153,983$ males and $2,160,976$ females surviving individuals at age 20 in the first generation.

Note that we compute the EPVs for males and females separately, see Section 8.3.1.

(b) EPVs at age 20 in the HCM families are estimated by the Monte Carlo simulation, as described in Section 8.3.2, using 500 simulations (which is sufficient to meet the actuarial equivalence principle, see the following sections). Consequently, when quantile intervals are cited in any results, they relate only to uncertainty arising from the simulation in respect of the HCM families. We represent 95% quantile intervals (QI, 95%) of 500 simulations. Note that we refer to QI not confidence intervals (CI) because of the small adjustment to premium rates made in Section 8.6.2.

(c) Note that the simulated total losses at age 20 in the HCM families, see $\left(\sum_{m=1}^{\Omega_e} L_{m,e}\right)$ in equation (8.6), form a set of identically distributed and independent random variables. From the Central Limit Theorem, the Monte-Carlo sampling distribution of aggregated losses, $E[L^{\mathcal{N}}] + L^{\mathcal{H}}$, is asymptomatically normal, with

(i) mean 0 if there is no adverse selection,

(ii) mean>0 if there is adverse selection (in our study),

and some variance $\sigma^2_{\mathcal{E}}$ depending on $\mathcal{E}$.

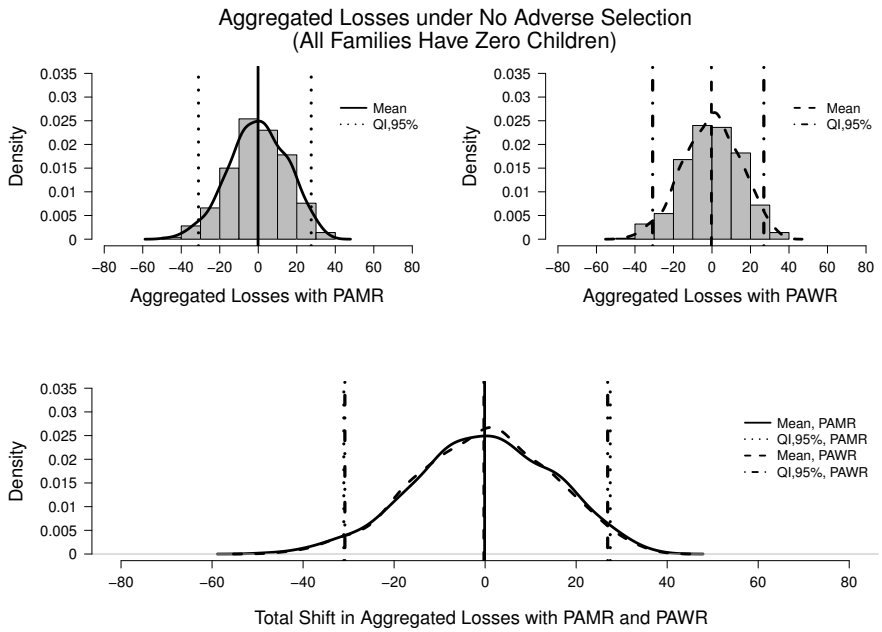


Figure 8.1: Distribution of aggregated losses at age 20, $E[L^{\mathcal{N}}] + L^{\mathcal{H}}$, where all families have zero children, under no adverse selection with the baseline assumptions in the whole population. PAMR: Premiums Associated with Mortality Rates. PAWR: Premiums Associated with Weighted Average Rates Over All Nine Model Sub-populations.

8.5 Baseline Assumptions

Table 8.1 summarises all the baseline assumptions in the adverse selection model (Chapter 7). Many of these assumptions were shown in the earlier chapters (Tables 3.10, 5.1, 7.1, and 7.2). Our baseline $\lambda = 1.8$ for the average number of children per family was mentioned in Section 5.6.1, but it was not shown in any table before, therefore it was added here.

We pointed out before, we numerically solve the Kolmogorov forward equations with time step 0.0005 years (Sections 3.11, 5.2.3, and 7.4.5) which is the same for the numerical solution of Thiele's equations. And, we simulate the lives in HCM families by using time step 0.005 years (Sections 5.2.3 and 7.6) for the reason in the following Section 8.6.3 in this chapter (See also Figure 8.4).

8.6 Aggregated Losses: No Adverse Selection

In Figure 8.1, we show the expected present value of aggregated losses at age 20 under no adverse selection in the zeroth generation only, which means neither HCM

Table 8.1: Baseline assumptions for the adverse selection model parameters.

Epidemiological Parameters	Table 3.10	Section 3.10
Prevalence of non-HCM mutations in the general population at age 20	0.998	Section 3.7.1
Prevalence of HCM mutations in the general population at age 20	0.002	Section 3.7.1
Prevalence of known early-onset mutations in the HCM population at birth	0.5	Section 3.7.2
Prevalence of known late-onset mutations in the HCM population at birth	0.1667	Section 3.7.2
Prevalence of unknown early-onset mutations in the HCM population at birth	0.25	Section 3.7.2
Prevalence of unknown late-onset mutations in the HCM population at birth	0.0833	Section 3.7.2
Penetrance of early-onset HCM at age 20	100%	Section 3.8.1
Penetrance of late-onset HCM at ages 20–70	Figure 3.4	Section 3.8.2
Hazard rate of fatal HCM per annum for all ages	0.0055	Section 3.9.7
Hazard rate of non-fatal HCM per annum for all ages	Table 3.6	Section 3.9.7
Hazard rate of all other death per annum for all ages	Figure 3.9	Section 3.10
Hazard rate of testing in one year at ages 0–70 if proband exists in family	0.6931472	Section 4.3.2
Average number of children per family (λ)	1.8	Section 5.6.1
Premium Rates	Equation (7.1)	Section 7.4.3
Benefit (Sum Assured)	£1	Section 7.4.5
Force of interest per annum	0.05	Section 7.4.5
Hazard rate of normal purchase per annum at ages 20–60	0.05	Section 7.5.1
Hazard rate of purchase per annum at ages 20–60 based on information class of individuals	Table 7.2	Section 7.5.1

nor non-HCM families have children. In Figure 8.1:

- The upper left plot represents the histogram (of 500 simulations) of the expected present value of aggregated losses at age 20 where each insured person, as long as they are alive during the policy term, is charged premium rates associated with true mortality rates (PAMR) under no adverse selection (see Section 6.5). This assumes that the insurer knows which sub-population each individual is in and charges the correct mortality rate as the premium rate (associated with the purchased £1 sum assured). Of course this is impossible, it simply serves as a check that the mean loss is zero (or if the actuarial equivalence principle is satisfied).

- As a second check, we assume there is no adverse selection and calculate premium rates as weighted average rates over all nine model sub-populations (PAWR), see equation (7.1). This assumes that the insurers do not know which sub-population any individual is in. See Appendix A.1. Therefore, the upper right plot represents the histogram (of 500 simulations) of the expected present value of aggregated losses at age 20 where each insured person, as long as they are alive during the policy term, is charged these weighted average premium rates (PAWR).

- The upper plots also present a kernel density estimate, mean and 95% quantile intervals (QI, 95%). These plots are consistent with point (c) (i) in Section 8.4.

- The bottom plot compares the upper plots.

In Figure 8.2, we show the expected present value of aggregated losses at age 20, under no adverse selection, including the zeroth and first generations, which means we allow families to have children. In Figure 8.2:

- The upper left plot represents the histogram (of 500 simulations) of the expected present value of aggregated losses at age 20 where each insured person, as long as they are alive during the policy term, is charged the PAMR premium rates.

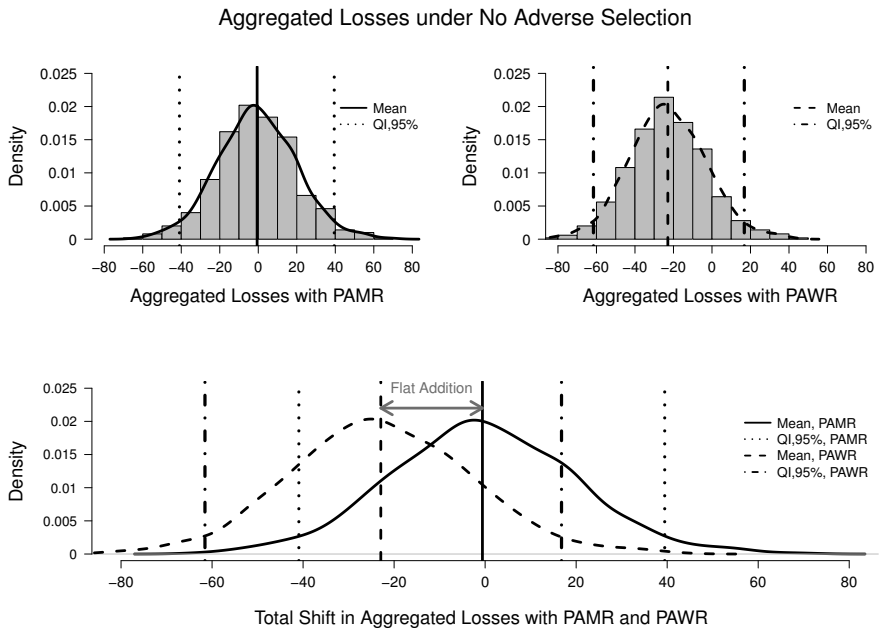


Figure 8.2: Distribution of aggregated losses at age 20, $E[L^{\mathcal{N}}] + L^{\mathcal{H}}$, under no adverse selection with the baseline assumptions in the whole population. PAMR: Premiums Associated with Mortality Rates. PAWR: Premiums Associated with Weighted Average Rates Over All Nine Model Sub-populations.

- The upper right plot represents the histogram (of 500 simulations) of the expected present value of aggregated losses at age 20 where each insured person, as long as they are alive during the policy term, is charged the PAWR premium rates.

- The upper plots also present a kernel density estimate, mean and 95% quantile intervals (QI, 95%). While the upper left plot is consistent with point (c) (i) in Section 8.4, the upper right plot is not (it is negatively shifted). The reason for this discrepancy is explained in Section 8.6.1.

- The bottom plot compares the upper plots and shows as a flat addition the difference between two approaches of charging premium rates, PAMR and PAWR.

8.6.1 Prevalences of Gene Mutations Over Time

In the adverse selection model, with the baseline assumptions in Table 8.1 , the prevalence rate of HCM-related mutations is 0.00226 at birth in the zeroth generation. See Table 5.2. (Note that the prevalence 0.00226 represents the prevalence of

HCM mutations at calendar time 0 in the general population in the testing model, see Section 5.3.3. This prevalence is also same for the adverse selection model because nobody purchases insurance at calendar times 0–20). However, when families have children at calendar time 30, we observe the prevalence of HCM-related mutations at birth in the first generation to be less than 0.00226. This is because some of the mutation carriers in the zeroth generation have died before age 30. So, if we charge children premium rates based on mutation prevalence 0.00226, we charge them too much. This 'unexpected' income gives a negative shift to the distribution of the expected present value of aggregated losses at age 20.

This result has interesting consequences in our model.

- We proportionally have fewer individuals carrying HCM mutations in the first generation compared to the zeroth generation.
- The prevalence of HCM-related mutations decreases over generations because HCM-related mortality is quite high before the assumed reproductive age 30.
- It implies that if we run our simulation through many generations, then HCM-related mutations would disappear over time.

The genetic literature (see Falconer & Mackay (1990) and Sudbery (2002)) discusses this phenomenon as follows:

- The replacement of 'unfavourable' gene mutations with 'favourable' ones over time is called 'selection'.
- However, 'selection' is not the only factor altering prevalences of gene mutations over generations. New mutations might arise in unmutated genes as long as selection then acts to eliminate them.
- There is an inverse relationship between 'mutation' and 'selection' which keeps prevalences of gene mutations in balance over generations. It would be a significant factor if we had modelled more generations in this study. For simplicity, we fix this issue with an adjustment, which is a constant premium factor appearing in Section 8.6.2.

Table 8.2: Estimated value of $E[L]$, the EPV of aggregated losses at age 20 in the whole population, see equation (8.2) where the model uses different premium rates approaches with the baseline assumptions. PAMR: Premiums Associated with Mortality Rates. PAWR: Premiums Associated with Weighted Average Rates Over All Nine Model Sub-populations. aPAWR: Adjusted PAWR. The numerics below are rounded to six decimal points.

Premium Rates	Estimated EPV (Aggregated Losses)
PAMR	-0.591222
PAWR	-22.923866
aPAWR	-0.058020

- There are, moreover, other factors changing prevalences of gene mutations, such as random drift (random changes affecting prevalences of gene mutations) and immigration which can both be significant in small populations. They would not be significant factors for our model.

8.6.2 Premium Factor and Adjusted Premium Rates

Due to the reasons in Section 8.6.1, the estimated value of $E[L]$ is negative (not zero), see the upper right plot in Figure 8.2. We define a constant premium factor Π, satisfying the condition in point (c) (i) in Section 8.4, obtained by the fraction of the estimated value of EPV of aggregated future benefit at age 20, which is computed by the combination of equations (8.5) and (8.8), over the estimated value of EPV of aggregated future income at age 20, which is computed by the combination of equations (8.4) and (8.7), under charging children higher premium rates.

We find, with the baseline assumptions in Table 8.1, $\Pi = 0.999859$. Table 8.2 shows the impact of the factor Π by comparing the estimated values of $E[L]$ of three kinds:

(a) the model uses the PAMR premium rates,

(b) the model uses the PAWR premium rates,

(c) the model uses the adjusted PAWR premium rates by the factor Π (for all individuals) referring to this as aPAWR.

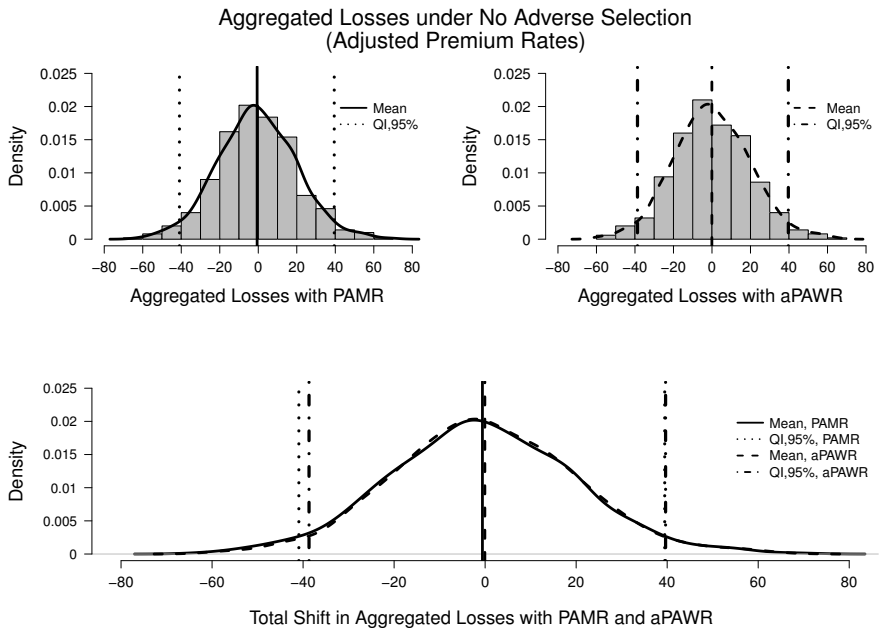


Figure 8.3: Distribution of aggregated losses at age 20, $E[L^{\mathcal{N}}] + L^{\mathcal{H}}$, under no adverse selection with the baseline assumptions and adjusted premium rates in the whole population. PAMR: Premiums Associated with Mortality Rates. aPAWR: Adjusted Premiums Associated with Weighted Average Rates Over All Nine Model Sub-populations.

Note that from now on the results in this chapter or the following chapters we will base our model on the aPAWR premium rates.

In Figure 8.3, we show the expected present value of aggregated losses at age 20 under no adverse selection with the aPAWR premium rates. In Figure 8.3:

- The upper right plot shows the histogram (of 500 simulations) of the expected present value of aggregated losses at age 20 under no adverse selection with baseline assumptions in the whole population where we charge individuals the aPAWR premium rates.
- The upper left plot is the same as that in Figure 8.2.
- The upper plots also present a kernel density estimate, mean and 95% quantile intervals (QI, 95%).
- The bottom plot compares the upper plots, which are both consistent with point (c) (i) in Section 8.4.

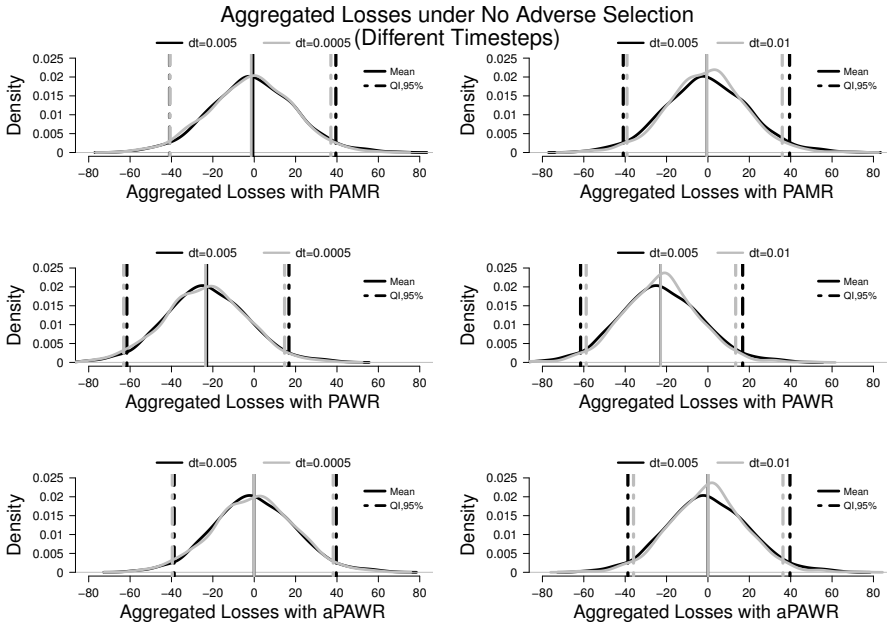


Figure 8.4: Distribution of aggregated losses at age 20, $E[L^{\mathcal{N}}]+L^{\mathcal{H}}$, under no adverse selection with the baseline assumptions with different time step dt in the Monte Carlo simulation. PAMR: Premiums Associated with Mortality Rates. PAWR: Premiums Associated with Weighted Average Rates Over All Nine Model Sub-populations. aPAWR: Adjusted PAWR.

8.6.3 Different Time Steps

In Sections 3.11, 5.2.3, 7.4.5, and 8.5, we mentioned that we always solve the Kolmogorov or Thiele's equations with time step $dt = 0.0005$ in this study.

In Sections 5.2.3, 7.6, and 8.5, we mentioned that we simulate HCM families' life histories with time step $dt = 0.005$ because the expected present value of aggregated insurance losses with $dt = 0.005$ is sufficiently accurate compared with $dt = 0.0005$ shown in Figure 8.4.

8.7 Aggregated Losses under Adverse Selection

In Figure 8.5, we present kernel density estimates with 95% quantile intervals (QI, 95%) of the histograms (of 500 simulations) of the expected present value of aggregated losses at age 20 under different levels of adverse selection with the baseline assumptions where family history is not known to insurers, and where family history is known to insurers in Figure 8.6. In both figures:

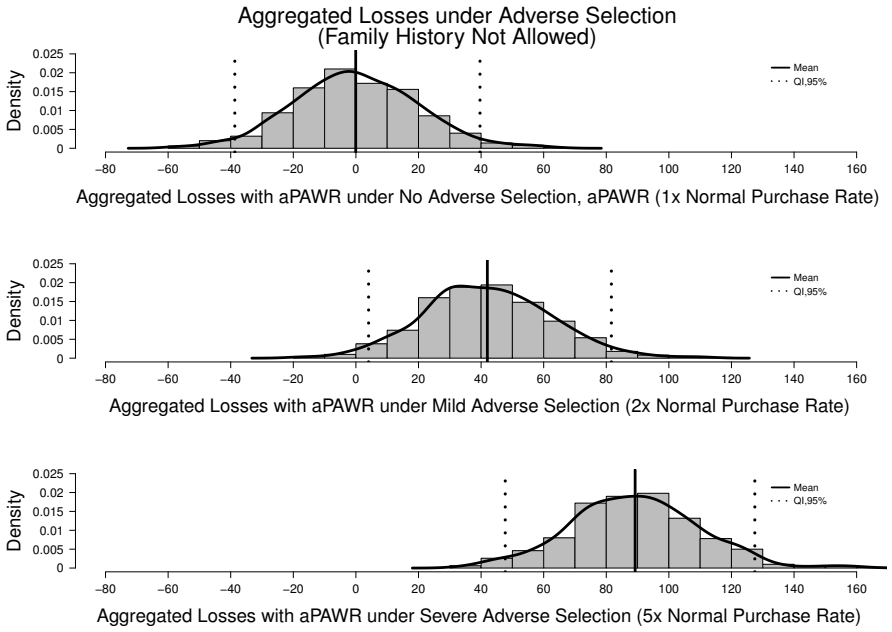


Figure 8.5: Distribution of aggregated losses at age 20 under different levels of adverse selection where family history is not disclosed to insurers with the baseline assumptions in the whole population. aPAWR: Adjusted Premiums Associated with Weighted Average Rates Over all Nine Model Sub-populations.

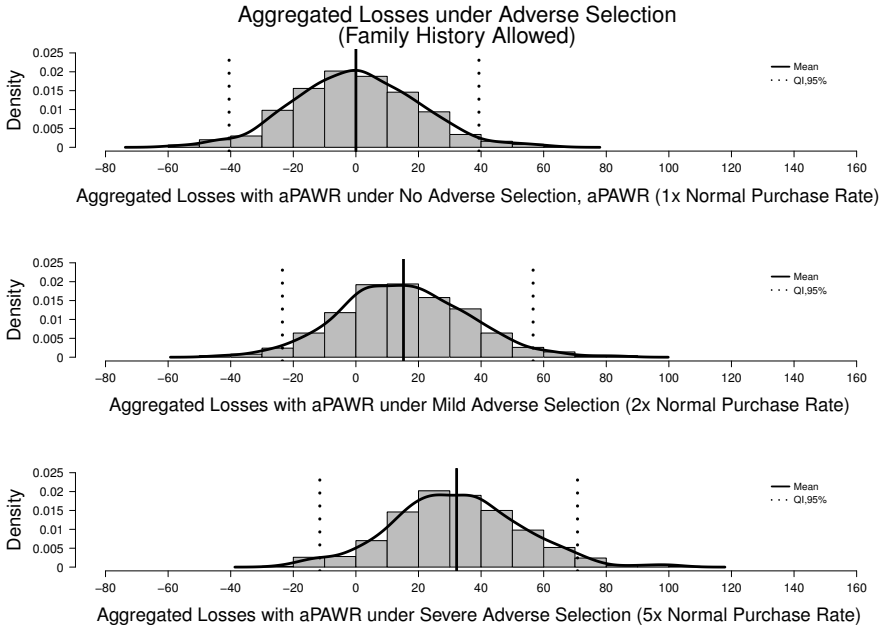


Figure 8.6: Distribution of aggregated losses at age 20 under different levels of adverse selection where family history is disclosed to insurers with the baseline assumptions in the whole population. aPAWR: Adjusted Premiums Associated with Weighted Average Rates Over All Nine/Eight Model Sub-populations.

(a) The top plots show the expected present value of aggregated losses at age 20 under no adverse selection which is consisted with point (c) (i) in Section 8.4. Note that when family history is known to insurers and an individual purchasing insurance with an appearing proband in the family:

  (i) individual should be charged weighted average premium rates over all *eight* sub-populations (in HCM families) in Figure 7.4.

  (ii) However, as we stated before, (see Sections 1.1 and 7.4.1), we assume negative test results are disclosed to insurers based on the UK case.

  (iii) Therefore, the individuals with negative test results in our model is, as a baseline assumption, charged weighted average premium rates over all *nine* sub-populations in Figure 7.4.

  (iv) Doing so under no adverse selection should have lead the distribution of aggregated losses to have mean>0, which is not consistent with Figure 8.6. This is not happening because we use the aPAWR premium rates as a baseline (Section 8.6.2). Note that this is conservative for our purposes.

(b) The middle and bottom plots shows the expected present value of aggregated losses at age 20 under mild and severe adverse selection which are consistent with point (c) (ii) in Section 8.4. See Table 7.2 for mild and severe purchase scenarios under adverse selection.

(c) We observe the losses in both figures under no adverse selection are approximately normally distributed around zero. However, with adverse selection, the estimated mean losses in Figure 8.5 are approximately 2.7 times of those in Figure 8.6.

8.8 A Measure of Adverse Selection Costs

We discussed through the thesis that we want to model changing behaviour of individuals under adverse selection and calculate the associated adverse selection costs.

In this section, we summarise the discussion on adverse selection and introduce a measure of its costs in our model.

Adverse selection by definition is an 'asymmetry' of access to information between insurers and individuals. In our model, individuals always have 'more information' about their risks than insurers do under adverse selection because either: genetic test results are not disclosed to the insurers, but family history is, or; both genetic test results and family history are not disclosed to insurers. (Note that negative test results are assumed to be disclosed in our model, see Sections 1.1 and 7.4.1).

Therefore, we assume in the presence of adverse selection (note that in the absence of adverse selection the expected present value of aggregated losses at age 20 is $E[L] = E\left[L^{\mathcal{H}} + L^{\mathcal{N}}\right] = 0$, see equation (8.2)):

(a) The adverse selectors might purchase insurance at a higher rate with the same benefit as non-adverse selectors. Under adverse selection, the expected present value of aggregated losses at age 20 with the same sum assured purchased by all is:

$$E\left[L^{*\mathcal{H}} + L^{\mathcal{N}}\right] > 0, \tag{8.9}$$

where $E\left[L^{*\mathcal{H}}\right]$ refers to the expected present value of aggregated losses at age 20 from HCM families under adverse selection. Following Section 6.7 and Appendix A.1.3, we measure the adverse selection costs as follows:

$$\frac{E\left[L^{*\mathcal{H}} + L^{\mathcal{N}}\right]}{E\left[I^{*\mathcal{H}} + I^{\mathcal{N}}\right]} \tag{8.10}$$

where $E\left[I^{*\mathcal{H}}\right]$ refers to the expected present value of aggregated premium income at age 20 from HCM families under adverse selection. This ratio can be interpreted as the proportionate increase in all premiums that would be necessary to recoup the cost of adverse selection.

(b) The adverse selectors might purchase insurance at a higher rate with a higher benefit. Then, decompose $L^{*\mathcal{H}}$ into two pieces: $L^{*\mathcal{H}(1)}$, representing the losses from adverse purchasers (higher rate and higher benefit) in HCM families,

and $L^{*\mathcal{H}(2)}$, representing the losses from normal purchasers (normal rate and normal benefit) in HCM families. Apply the same measure in equation (8.10):

$$\frac{E\left[\alpha L^{*\mathcal{H}(1)} + L^{*\mathcal{H}(2)} + L^{\mathcal{N}}\right]}{E\left[\alpha I^{*\mathcal{H}(1)} + I^{*\mathcal{H}(2)} + I^{\mathcal{N}}\right]} \tag{8.11}$$

where the individuals contributing to $L^{*\mathcal{H}(1)}$ purchase insurance of amount α times the normal sum assured £1.

Chapter 9

An Adverse Selection Model of Hypertrophic Cardiomyopathy (HCM) for Life Insurance III: Results

9.1 Introduction

Chapter 7 presented the methodology of insurance loss calculations in respect of an individual in the adverse selection model (Figure 7.1). Chapter 8 extended the methodology to the whole population with a measure, see Section 8.8, capturing the adverse selection costs in terms of increased premiums.

This chapter gives the results in respect of our measure in equation (8.10) for £1 sum assured or equation (8.11) for £α sum assured. They are the mean premium increases (to redeem the mean adverse selection costs) with the 95% quantile intervals (QI, 95%) obtained from 500 simulations of the HCM families where adverse genetic test results are not disclosed to insurers (negative test results are disclosed to insurers in our study based on the moratorium in the UK, see Sections 1.1 and 7.4.1). Family history may or may not be disclosed to insurers. Section 9.2 shows the results under the baseline scenario. Sections 9.3; 9.4; 9.5, and 9.6 show the results associated with changes in the epidemiological; genetic testing; pricing; and

behavioural parameters in the baseline scenario. Section 9.7 discusses the results in terms of which factors (or the model parameters) significantly amplify and diminish the adverse selection costs.

9.2 Baseline Scenario, *sc.*0

Label *sc.*0 in Table 9.1 presents the mean premium increases in the baseline scenario where all the adverse selection model parameters have the values given in Table 8.1.

(a) The mean premium increases seem very low (especially when family history is allowed to insurers, it is statistically significant to state that insurers do not suffer from adverse selection under the baseline scenario) even under severe adverse selection. They are fractions of one percent.

(b) When family history is not known to insurers, the mean premium increases are about 2.7 times of those when family history is known to insurers.

Also, in Table 9.2, we present the contributions of the different sub-populations in HCM families (Figure 7.3) to the mean premium increases under adverse selection in the baseline scenario. With family history not allowed, only 0.9775% and 0.8644% of the premium increases under 'mild' and 'severe' adverse selection respectively are explained by the late-onset mutations. With family history allowed, late-onset mutations actually reduce premium increases, and the corresponding contributions are −6.9075% and −7.2983%. The prevalences of early- and late-onset mutations are clearly significant for the adverse selection costs, a point that we discuss further in Section 9.3.3.

9.2.1 A Note on the Decimal Places of the Results

All the results represented in the baseline scenario or the scenarios in the following sections are rounded to six decimal places; in the underlying calculations, although they are represented as some percentage to four decimal places.

The mean premium increases under no adverse selection in almost all scenarios are zero percent to four decimal places due to the small adjustment of premium

rates described in detail in Section 8.6.2. Note also that in a couple of scenarios, the mean premium increases under no adverse selection were -0.0001%, zero percent to three decimal places and they were rounded to be 0.0000%, though they were still zero percent to four decimal places before they were rounded to six decimal places.

9.3 Epidemiology

This section surveys what epidemiological parameters give significant increases or decreases to the mean premium increases under adverse selection.

Table 9.3 presents the mean premium increases associated with changes in the epidemiological parameters in the baseline scenario, *sc.*0, see Section 9.2.

- Section 9.3.1 describes the scenario labelled by *sc.*1 in Table 9.3.
- Section 9.3.2 describes the scenario labelled by *sc.*2 in Table 9.3.
- Section 9.3.3.2 describes the scenario labelled by *sc.*3 in Table 9.3.
- Section 9.3.4 describes the scenario labelled by *sc.*4 in Table 9.3.

9.3.1 Higher Mutation Prevalences: *sc.*1

In respect of Sections 2.5 and 3.7:

(a) The prevalence of clinical HCM in the general population, estimated to be 0.2% in Maron et al. (1995), was assumed (as a baseline) to be the prevalence of HCM-related mutations in the general population in this study. Doing so was conservative for our purposes. Howard (2014) also assumed (as a baseline) that the prevalence of HCM-related mutations is 0.2% in the general population, see Section 1.3.

(b) However, Bick et al. (2012) estimated the prevalence of known HCM-related mutations (0.6%) to be about three times higher than the prevalence of clinical HCM (0.2%) in the general population. Extrapolating from Bick et al. (2012), including for unknown mutations (see Section 3.7.2), the prevalence of HCM-related mutations is estimated as 0.9% in this study.

Table 9.1: Percentage increases in premium rates associated with the baseline scenario. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages. The 95% quantile intervals are in respect of the Monte-Carlo simulation of mean EPVs of cashflows in the HCM families.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.0	0.2%	1	F^{b**}_x	0.6931472	$\rho^{nf\ddagger}_x$	0.0055	1.8	5%;£1	5%;£1	None	%0.0000	$(-0.0239, 0.0245)$	%0.0000	$(-0.0250, 0.0242)$
								5%;£1	10%;£1	Mild	0.0259	$(0.0025, 0.0503)$	0.0094	$(-0.0145, 0.0349)$
								5%;£1	25%;£1	Severe	0.0550	$(0.0294, 0.0786)$	0.0198	$(-0.0071, 0.0436)$

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.

Table 9.2: Contributions of different sub-populations to the mean increases in premium rates in the baseline scenario, *sc.*0, see Table 9.1.

Adverse		Family History: ✗			Family History: ✓		
Selection	Mutation	Early-Onset	Late-Onset	Total	Early-Onset	Late-Onset	Total
Mild	Known	%65.9932	%0.6146	%66.6078	%83.6378	%−4.0889	%79.5489
	Unknown	33.0293	0.3629	33.3922	23.2697	−2.8186	20.4511
	Total	99.0225	0.9775	100.0000	106.9075	−6.9075	100.0000
Severe	Known	66.2341	0.5579	66.7920	83.9440	−4.3126	79.6314
	Unknown	32.9015	0.3065	33.2080	23.3543	−2.9857	20.3686
	Total	99.1356	0.8644	100.0000	107.2983	−7.2983	100.0000

(c) From points (a) and (b) above, if clinical HCM is associated with 2/9 of HCM-related mutations is 0.2%, something is causing the remaining 7/9 'silent' HCM-related mutations. If we are to reproduce the observed incidence of HCM-related events from selected populations in which clinical HCM is present, then the penetrance of HCM might have been overestimated in the past. This would lead the mean premium increases to be reduced under adverse selection if either positive test result or family history is a strong incentive to purchase more insurance since the number of adverse selectors is larger, but the number of the individuals who are clinically affected by HCM does not change.

(d) Moreover, as Maron et al. (2016b) and Husser et al. (2018) showed not everyone with clinical HCM is diagnosed as having HCM. These studies pointed out that the prevalence of clinically-diagnosed HCM lies between 0.035–0.07%. If we had modelled the onset of clinical HCM as a transition into a state, this would reduce the adverse selection costs, since the early-onset mutation carriers surviving up to age 20 are assumed all to have clinical HCM at age 20 ($F(20) = 1$, see Section 3.8.1).

Since the hazard rates of HCM-related events are conditional on the presence of clinical HCM in the epidemiological literature, we can attribute the difference in prevalence to reduced penetrance.

As a result, to obtain the same clinical outcomes with the mutation prevalence of 0.9% as we had with the baseline assumptions (Table 8.1), we should adjust $F(x)$, the penetrance of HCM, see Sections 3.3.3 and 3.8. Doing so is not trivial because $F(x)$ is a function depending on age x. Our approach is simply to multiply $F(x)$ by a constant factor of 22% ($\approx$0.002/0.009) because we have no basis for any more sophisticated approach. This is crude (but no more so than the conservative assumption that $F(20) = 1$ for early-onset mutations).

The results are shown in label *sc.*1 in Table 9.3. The adverse selection costs diminish by a factor of about three. The prevalence of 'silent' HCM mutations could be a significant factor for the adverse selection costs.

9.3.2 The Penetrance of Late-Onset HCM: *sc.*2

We replace Christiaans et al. (2011) (Figure 3.4) with Terauchi et al. (2015) (Figure 3.5) for the penetrance of late-onset HCM (Section 3.8.2).

The results are shown in label *sc.*2 in Table 9.3. The adverse selection costs are not significantly affected because of the small proportion of the contributions of late-onset mutation carriers to the adverse selection costs, see Table 9.2.

9.3.3 The Proportions of HCM-Related Mutations

9.3.3.1 Late-Onset Mutations

We assume that the late-onset mutations account for 25% of mutations (see Table 3.2). From Table 9.2, we observe that the contributions of late-onset mutations to the adverse selection costs are very small. The 25% was extrapolated to unknown mutations. If doing this should be an underestimate, the adverse selection costs would reduce. Note that we do not model this.

9.3.3.2 Known Mutations: *sc.*3

We also assume (see Table 3.2) that the known mutations, ignoring unrelated mutations, account for 2/3 (baseline) to 3/4 (sensitivity) of the mutations.

(a) We presented the results in the baseline scenario, *sc.*0, see Table 9.1, when the known mutations account for 2/3 of the mutations.

(b) This section shows the results when the known mutations account for 3/4 of the mutations in the baseline scenario, *sc.*0, see Table 9.1.

The results are shown in label *sc.*3 in Table 9.3. This adjustment in the proportions of known and unknown mutations seems to be not significant for the adverse selection costs. We note the baseline assumption of the same purchase behaviour of individuals in information classes ζ^{50} and ζ^{100} (see Table 7.2) (note that this was conservative for our purposes). We will re-consider this analysis when we change the assumption for the purchasing behaviour of the individuals in information class ζ^{50} in Section 9.6.

9.3.4 Higher HCM-Related Mortality: *sc.*4

In Section 3.9, we discussed the hazard rates of HCM-related endpoints, in particular the hazard rate of fatal HCM. We also noted that an annual mortality rate of HCM of $q_x = 0.01$ was widely cited in the literature, see Section 1.5. This was used in Howard (2014), see Sections 1.3 and 3.9.8, and is considerably higher than our estimated annual hazard rate of 0.0055. Therefore, similar to Section 9.3.1, we make a combined assumption by replacing the baseline hazard rate of fatal HCM (0.0055 per annum at all ages) with the sensitivity hazard rate of fatal HCM (0.01 per annum at all ages), and setting the hazard rate of non-fatal HCM (Table 3.6) to be zero at all ages.

The results are shown in label *sc.*4 in Table 9.3. The mean premium increases are only slightly higher even though we almost doubled the hazard rate of fatal HCM. This happens either because of our assumption that the hazard rate of non-fatal HCM is zero while the hazard rate of fatal HCM is 0.01 per annum at all ages, or the adverse selection costs are dominated by the other factors (or the other parameters) of the model. Or, the insurer knows and uses the 'true' fatal HCM hazard rate in calculating C^1 (see Section 7.4.4) premium rates.

We will consider the other factors of the model such as genetic testing and pricing in the following sections. However, this epidemiological parameter is quite important. Partly this is because a life insurance study, and insurers pay benefits to the policyholders on the event of death. It is partly because Macdonald & Yu (2011) and Howard (2014) presumably agreed on the majority of difference between the adverse selection costs reported in both studies was caused by cardiomyopathies because these disorders showed higher mortality compared to the other disorders in Howard (2014), see Section 1.1. It is partly because this epidemiological parameter is subject to selection bias, see point (e) in Section 9.7.2. Therefore, we will probe the impact of this epidemiological parameter on the adverse selection costs, by doing a sensitivity analysis in Section 9.7.4.

Table 9.3: Percentage increases in premium rates associated with changes in the epidemiological parameters in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.1	**0.9%***	**0.22**	$\mathbf{0.22}\boldsymbol{F^{b}_{x}}$**	0.6931472	ρ^{nf}_x‡	0.0055	1.8	5%;£1	5%;£1	None	%0.0000	(−0.0394, 0.0391)	%0.0000	(−0.0392, 0.0388)
								5%;£1	10%;£1	Mild	0.0090	(−0.0300, 0.0479)	0.0029	(−0.0358, 0.0419)
								5%;£1	25%;£1	Severe	0.0189	(−0.0205, 0.0582)	0.0062	(−0.0324, 0.0453)
sc.2	0.2%*	1	$\boldsymbol{F^{s}_{x}}$†	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	5%;£1	None	0.0000	(−0.0246, 0.0240)	0.0000	(−0.0246, 0.0235)
								5%;£1	10%;£1	Mild	0.0262	(0.0012,0.0509)	0.0082	(−0.0164, 0.0329)
								5%;£1	25%;£1	Severe	0.0560	(0.0294,0.0822)	0.0180	(−0.0086, 0.0441)
sc.3	**0.2%**#	1	F^{b}_{x}	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	5%;£1	None	0.0000	(−0.0226, 0.0254)	0.0000	(−0.0224, 0.0258)
								5%;£1	10%;£1	Mild	0.0259	(0.0012,0.0515)	0.0098	(−0.0150, 0.0353)
								5%;£1	25%;£1	Severe	0.0550	(0.0302,0.0802)	0.0207	(−0.0040, 0.0458)
sc.4	0.2%*	1	F^{b}_{x}	0.6931472	**0**	**0.01**	1.8	5%;£1	5%;£1	None	0.0000	(−0.0240, 0.0251)	0.0000	(−0.0236, 0.0249)
								5%;£1	10%;£1	Mild	0.0301	(0.0028,0.0565)	0.0095	(−0.0164, 0.0353)
								5%;£1	25%;£1	Severe	0.0643	(0.0338,0.0936)	0.0207	(−0.0087, 0.0484)

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
\# $p^{\mathcal{H}}$ uses sensitivity sub-population prevalences in the HCM population in Table 3.2.
** F^{b}_{x}: Baseline penetrance of late-onset HCM in Figure 3.4.
† F^{s}_{x}: Sensitivity penetrance of late-onset HCM in Figure 3.5.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.

9.4 More Genetic Testing

9.4.1 Higher Test Rate in Nuclear Families: *sc.*5

In Section 4.3.2, we denoted the annual hazard rate of the uptake of the testing at-risk relatives after an appearing proband in the family by $\mu_{x,z}$ where x is age and z is duration in years in an untested state since the appearance of a proband in the family. See equation (4.1) showing we assume (as a baseline) $\mu_{x,z} = 0.6931472$ where $0 < z \leq 1$, equivalent to 50% of at-risk relatives being tested in one year after a proband appearing in the family; otherwise, zero. We consider in this section an higher rate of $\mu_{x,z} = 4.60517$ where $0 < z \leq 1$, equivalent to 99% of at-risk relatives being tested in one year after a proband appearing in the family; otherwise, zero.

The results are shown in label *sc.*5 in Table 9.4.

(a) When family history is not disclosed to insurers, the adverse selection costs are almost identical since our baseline assumption of that the individuals in information class ζ^{50} behave, in terms of purchasing insurance, in the same way as those in information class ζ^{100} (see Table 7.2). This assumption is relaxed in Section 9.6.

(b) When family history is disclosed to insurers:

  (i) The adverse selection costs are higher because we calculate premium rates as if negative genetic test results are not disclosed to the insurers (see point (a) in Section 8.7). But as we follow the UK case in which negative test results are disclosed to the insurers, see Sections 1.1 and 7.4.1, negatively tested individuals in our model are charged less than they are supposed to be when family history is known to insurers.

  (ii) As a result, the increased hazard rate of the uptake of genetic testing brings more negatively tested individuals giving the adverse selection costs to increase.

  (iii) Note that the increased hazard rate of genetic testing brings more positively tested individuals, as well. However, since we keep our baseline

assumption for the purchase behaviour of information class ζ^{50}, see point (a) above, we would not expect to see any impact of more positively tested individuals on the adverse selection costs.

9.4.2 The Extension of Testing Beyond Nuclear Families: $sc.6 - sc.8$

We noted, in Sections 5.1 and 5.3.1, that this study models only nuclear families. However, we can increase λ, the average number of children per family, see Section 5.6.1, as a proxy of cascade genetic testing (Section 4.2) spreading beyond the nuclear family. This means that a single proband exposes more at-risk relatives to take up genetic testing. In this case, each family has a larger number of children (the first generation), however, the results are still reasonable since the nature of cascade genetic testing implies any relative before taking up the testing is known to be a mutation carrier with 1/2 probability.

The results are shown in labels *sc.*6, *sc.*7, and *sc.*8 for $\lambda = 3.0, 5.0$, and 7.0, respectively in Table 9.4. The last of these increases the adverse selection costs by a factor of 2.5. This is an approximation of 'cascading' genetic testing from the proband's family into roughly three other related nuclear families.

Moreover, it is fair to note that more genetic testing would bring more individuals into treatment for HCM, and, if effective, this would reduce mortality. See point (f) in Section 9.7.2.

9.5 Pricing

This section surveys what pricing parameters give significant increases or decreases to the mean premium increases under adverse selection.

Table 9.5 presents the mean premium increases with changes in the pricing parameters in the baseline scenario, *sc.*0, see Section 9.2.

- Section 9.5.1 describes the scenario labelled by *sc.*9 in Table 9.5.
- Section 9.5.2 describes the scenarios labelled by $sc.10 - sc.12$ in Table 9.5.

Table 9.4: Percentage increases in premium rates associated with changes in the model parameters resulting more genetic testing in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages. Note that the mean premium increases in respect to $\lambda > 1.8$ which is a proxy of cascade genetic testing spreading outside the nuclear family (Section 9.4.2).

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.5	0.2%	1	F^{b**}_x	**4.60517**	$\rho^{nf\ddagger}_x$	0.0055	1.8	5%;£1	5%;£1	None	%0.0000	$(-0.0240, 0.0245)$	%0.0000	$(-0.0247, 0.0237)$
								5%;£1	10%;£1	Mild	0.0259	(0.0022,0.0498)	0.0131	$(-0.0110, 0.0377)$
								5%;£1	25%;£1	Severe	0.0551	(0.0297,0.0779)	0.0276	(0.0017,0.0507)
sc.6	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	**3.0**	5%;£1	5%;£1	None	0.0000	$(-0.0196, 0.0201)$	0.0000	$(-0.0198, 0.0195)$
								5%;£1	10%;£1	Mild	0.0372	(0.0153,0.0591)	0.0133	$(-0.0088, 0.0356)$
								5%;£1	25%;£1	Severe	0.0796	(0.0573,0.1033)	0.0288	(0.0069,0.0525)
sc.7	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	**5.0**	5%;£1	5%;£1	None	0.0000	$(-0.0179, 0.0185)$	0.0000	$(-0.0179, 0.0178)$
								5%;£1	10%;£1	Mild	0.0517	(0.0340,0.0704)	0.0186	$(0.0005, 0.0377)$
								5%;£1	25%;£1	Severe	0.1108	(0.0909,0.1320)	0.0404	(0.0212,0.0623)
sc.8	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	**7.0**	5%;£1	5%;£1	None	0.0000	$(-0.0147, 0.0148)$	0.0000	$(-0.0153, 0.0144)$
								5%;£1	10%;£1	Mild	0.0631	(0.0474,0.0799)	0.0228	(0.0075,0.0399)
								5%;£1	25%;£1	Severe	0.1349	(0.1160,0.1536)	0.0494	(0.0294,0.0672)

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.

9.5.1 A Smaller Life Insurance Market: *sc.*9

Our baseline assumption of normal purchase (hazard) rate per annum at ages 20–60 is 5%, representing a large life insurance market (Macdonald & Yu 2011) (Sections 7.4.5 and 7.5.1). Here we consider a smaller normal purchase (hazard) rate per annum at ages 20–60 of 1%, representing a smaller life insurance market.

The results are shown in label *sc.*9 in Table 9.5. We observe significant increases in the mean premium increases, approximately 1.6 and 6 times of those in the baseline scenario under mild and severe adverse selection. The smaller purchase rate causes the adverse selection costs to increase due to a smaller proportion of non-HCM families purchasing insurance.

A similar effect would have been achieved by having a higher purchase rate combined with a significant lapse rate. In Section 10.5, we model lapse with the assumed lapse rates in Howard (2014) (see Section 10.2.3).

9.5.2 Higher Sums Assured: *sc.*10 − *sc.*12

The adverse selection costs, in Gutiérrez & Macdonald (2004), were very nearly proportionate to any increase in the sum assured taken out by adverse selectors because the expected present value of the actuarial losses in each underwriting class defined there were zero under no adverse selection. Therefore, the effect of larger sums assured, when adverse selection arises in one or more of these underwriting classes, could be computed as a multiple of the results obtained with £1 sum assured.

Here we measure differently the impact of higher sums assured, see equation (8.11). In our case, it was not possible to partition states into fixed underwriting classes since the underwriting classes depend on the random appearance of a proband. Thus, for large α, the term $\alpha L^{*\mathcal{H}(1)}$ in the numerator in equation (8.11) where α represents higher sum assured and $L^{*\mathcal{H}(1)}$ represents the losses from adverse selectors, who purchase insurance at a higher rate, in HCM families, dominates the numerator in the equation. Note that the dominator in equation (8.11) is dominated by $I^{\mathcal{N}}$, representing the premium income from non-HCM families, the majority of the population.

We also assume that the individuals in information classes ζ^{100} and ζ^{50} who purchase insurance purchase the same benefit. (Howard (2014) assumed that the normal sum assured was $\$100,000$, and adverse selectors purchased sums assured of 10 times of the normal sum assured, $\$1,000,000$. See Sections 1.3 and 10.2.2).

The results are shown in labels *sc.*10, *sc.*11, and *sc.*12 in Table 9.5 in respect of purchase of sum assured £2, 4, and 10 in information classes ζ^{100} and ζ^{50} while the rest purchase the normal sum assured of £1. Due to the reasoning from equation (8.11), the mean premium increases are higher than in proportion to the sum assured.

9.6 Information Classes: *sc.*13 – *sc.*16

We assume (as a baseline) that the individuals in information class ζ^{50} would behave in the same way as those in information class ζ^{100}, which was conservative for our purposes. See Table 7.2. Here we relax this assumption and consider different behaviours of the individuals in information class ζ^{50}.

The results are shown in labels *sc.*13, *sc.*14, *sc.*15, and *sc.*16 in Table 9.6.

(a) When family history is not allowed to insurers, and where the individuals in information class ζ^{50} purchase at the normal rate under mild or severe adverse selection, the adverse selection costs diminish by a factor of about three. Some of the results in Table 9.6 are not applicable (n/a) because we assume there is no reason that the individuals in information class ζ^{50} would purchase at a lower rate than normal, when there is only one underwriting class but this might happen when family history is disclosed because the premium rates are increased, based on the family history available to insurers, which is considered in what follows.

(b) When family history is allowed to insurers, and where the individuals in information class ζ^{50} purchase at the normal rate under mild or severe adverse selection, the adverse selection costs diminish by a factor of about 1.6. If these individuals purchase at a lower rate than normal under mild or severe adverse selection, the adverse selection costs diminish more. The most eye-catching

Table 9.5: Percentage increases in premium rates associated with changes in the pricing parameters in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.9	0.2%	1	F^{b**}_x	0.6931472	$\rho^{nf\ddagger}_x$	0.0055	1.8	**1%**; £1	**1%**; £1	None	%0.0000	(−0.0411, 0.0411)	%0.0000	(−0.0416, 0.0408)
								1%; £1	**2%**; £1	Mild	0.0431	(−0.0005, 0.0837)	0.0149	(−0.0275, 0.0568)
								1%; £1	**25%**; £1	Severe	0.3225	(0.2636,0.3810)	0.1143	(0.0571,0.1708)
sc.10	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	5%;**£2**	None§	0.0555	(0.0254,0.0863)	0.0188	(−0.0117, 0.0488)
								5%;£1	10%;**£2**	Mild	0.1073	(0.0782,0.1405)	0.0375	(0.0076,0.0714)
								5%;£1	25%;**£2**	Severe	0.1655	(0.1323,0.2001)	0.0584	(0.0246,0.0922)
sc.11	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	5%;**£4**	None§	0.1665	(0.1220,0.2123)	0.0564	(0.0121,0.1023)
								5%;£1	10%;**£4**	Mild	0.2699	(0.2210,0.3213)	0.0937	(0.0483,0.1436)
								5%;£1	25%;**£4**	Severe	0.3859	(0.3325,0.4485)	0.1352	(0.0837,0.1957)
sc.12	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%; £1	5%; **£10**	None§	0.4981	(0.4062,0.6020)	0.1683	(0.0788,0.2628)
								5%; £1	10%; **£10**	Mild	0.7549	(0.6431,0.8653)	0.2604	(0.1518,0.3686)
								5%; £1	25%; **£10**	Severe	1.0429	(0.9080,1.1917)	0.3628	(0.2259,0.5113)

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.
§ Adverse selectors purchase at normal rate, but higher sums assured.

result is that the adverse selection costs is almost zero when these individuals in information class ζ^{50} do not purchase at all under mild adverse selection.

As a result, normal or decreasing purchase rate of the individuals in information class ζ^{50} under adverse selection seems to be a significant factor diminishing the adverse selection costs. Note that we do not consider different sums assured in respect of information classes ζ^{50} and ζ^{100} under adverse selection. Doing so was conservative for our purposes. We will return to this discussion on the purchased sum assured in information classes ζ^{50} and ζ^{100} in point (a) in Section 9.7.2.

9.7 Discussion

9.7.1 Factors Amplifying Adverse Selection Costs

We summarise here the significant factors amplifying adverse selection costs based on the results in the earlier sections in this chapter.

(a) **Unavailability of family history**. As we already noted that inability to use family history in underwriting amplifies adverse selection costs considerably, by a factor of about 2.7 in the baseline scenario, *sc*.0, Section 9.2. Though, the adverse selection costs are still very small.

(b) **More genetic testing**. We obtained more genetic testing in two directions: either a higher test rate in nuclear families (Section 9.4.1) or the same test rate beyond nuclear families (Section 9.4.2) (note that we do not model a higher test rate beyond nuclear families because of the reason in point (i) below).

  (i) The first of these does not amplify adverse selection costs because of the assumption of the same purchase behaviour of information class ζ^{50} as information class ζ^{100} (note that this assumption was conservative for our purposes),

  (ii) The second of these amplifies adverse selection costs considerably, by a factor of 2.5 when we approximate cascade genetic testing from the proband's family into roughly three other related nuclear families.

Table 9.6: Percentage increases in premium rates associated with changes in the purchase behaviour parameter based on the information classes in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5,$ or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2,$ or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5,$ or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum			%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}	ζ^{100}	Family History: ✗		Family History: ✓	
sc.13	0.2%*	1	F^{b**}_x	0.6931472	$\rho^{nf\ddagger}_x$	0.0055	1.8	5%;£1	5%;£1	**10%**; £1	%0.0086	$(-0.0144, 0.0330)$	%0.0057	$(-0.0178, 0.0306)$
								5%;£1	5%;£1	**25%**; £1	0.0182	$(-0.0069, 0.0431)$	0.0119	$(-0.0132, 0.0365)$
sc.14	0.2%*	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	**2.5%**; £1	**10%**; £1		n/a	0.0029	$(-0.0185, 0.0274)$
								5%;£1	**2.5%**; £1	**25%**; £1		n/a	0.0092	$(-0.0133, 0.0347)$
sc.15	0.2%*	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	**0%**; £1	**10%**; £1		n/a	0.0000	$(-0.0218, 0.0206)$
								5%;£1	**0%**; £1	**25%**; £1		n/a	0.0053	$(-0.0160, 0.0276)$
sc.16	**0.2%**#	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	5%;£1	5%;£1	**10%**; £1	0.0096	$(-0.0132, 0.0360)$	0.0063	$(-0.0170, 0.0331)$
								5%;£1	5%;£1	**25%**; £1	0.0203	$(-0.0033, 0.0463)$	0.0132	$(-0.0104, 0.0389)$

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
# $p^{\mathcal{H}}$ uses sensitivity sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.

(c) **A smaller life insurance market**. Replacing the 5% annual purchase rate, representing a large market, with a 1% annual purchase rate, representing a smaller market amplifies adverse selection costs considerably, by a factor of 1.6 and 6 under mild and severe adverse selection, respectively. See Section 9.5.1. We also noted that a similar effect can be obtained by modelling lapse with a significant lapse rate in a large market, see Section 10.5.

(d) **Higher sums assured**. Adverse selectors insuring themselves at higher than the normal sum assured (£1) amplifies adverse selection costs, considerably. See Section 9.5.2.

(e) **Higher HCM-related mortality**. It is interesting to observe that increases in adverse selection costs in a large market are small when we almost doubled the annual hazard rate of HCM-related mortality. It might happen because we also made the annual hazard rate of non-fatal HCM zero at all ages. Or, it might happen the other model parameters (such as behavioural), in terms of purchasing insurance, are more influential than this epidemiological parameter. Or, the insurers use the 'true' fatal HCM hazard rate in calculating C^1 (see Section 7.4.4) premium rates. See Section 9.3.4. Thus, we do a sensitivity analysis to understand in a broader perspective of the impact of fatal HCM on adverse selection costs. See Section 9.7.4.

9.7.2 Factors Diminishing Adverse Selection Costs

In this section, we do two things: We summarise the factors diminishing adverse selection costs based on the results in the earlier sections in this chapter. We also consider other factors, diminishing adverse selection costs, not explicitly modelled that actuaries should be aware of.

(a) **Information classes**. We consider the purchase behaviour in information class ζ^{50}. Assuming (as a baseline) that the purchase behaviour of the individuals in information class ζ^{50} is the same as those in information class ζ^{100} was conservative for our purposes. In reality, we really do not know how the

individuals in information class ζ^{50} would behave in terms of purchasing insurance since they know only that they carry a mutation with 50% probability. Section 9.6 showed that a lower purchase rate of the individuals in information class ζ^{50} than those in information class ζ^{100} reduces adverse selections costs, considerably. Moreover, we did not consider a case such that the individuals in information class ζ^{50} purchase a lower sum assured than the individuals in information class ζ^{100}. For example, in Table 9.5, we assumed the individuals information class ζ^{50} purchase the same sum assured as those in information class ζ^{100} (this was our baseline for the sums assured in both information classes). See label *sc.*12 in Table 9.5 where the individuals in information classes ζ^{50} and ζ^{100} both purchase insurance at the same sum assured which is 10 times higher than the normal sum assured of £1. Doing so was also conservative for our purposes. In reality, would the individuals in information class ζ^{50} purchase at the very high sum assured (such as £10) since they only know that they have a 50% chance to carry the mutation? Would it be thought as a good financial investment for the individuals in information class ζ^{50}?

(b) **Higher mutation prevalences**. This topic is a central part of this study, since it emphasises that actuaries should pay attention to the biases of epidemiological data of two kinds, see points (i) and (ii) below, and Section 9.3.1. For example, in HCM,

(i) **Prevalence of 'silent' mutations**. Bick et al. (2012) estimated the prevalence of HCM-related mutations in the general population to be much higher than previously thought, based on a study by Maron et al. (1995), which for used on the prevalence of clinical HCM in the general population. The difference between these two studies show that there is something causing a significant proportion of the mutations to be 'silent'. And, modelling this diminishes adverse selection costs considerably, by a factor of about three. See point (c) in Section 9.3.1.

(ii) **Modelling a state of the onset of clinical HCM**. Recent studies (Maron et al. (2016b) and Husser et al. (2018)) relying on large databases

of reported cases of clinically affected HCM patients found that the prevalence of clinically-diagnosed HCM (0.035–0.07%) in the general population is much less than that of clinically-present HCM (0.2%). Modelling this would reduce the adverse selection costs since we assume $F(20) = 1$ for early-onset mutation carriers. See point (d) in Section 9.3.1.

(c) **The proportion of late-onset mutations**. This is related to our extrapolation for the unknown mutations relying on the proportion of known mutations. If our extrapolation is an underestimate, then we would observe decreases in adverse selection costs had we modelled them. See Section 9.3.3.1.

(d) **The manner of cascade testing for clinical HCM**. In our model, we assume that when a proband is tested and carries a known mutation, all at-risk relatives are offered genetic testing and may then purchase insurance. The model does not allow underwriters to acquire any other information. For example, after a proband appears in the applicant's family before purchasing insurance, clinical diagnosis of the applicant (or at-risk relatives) with HCM as a pre-existing condition by imaging machines (Section 2.2.2). This could also include symptomatic patients seeking insurance, without a clinical diagnosis of HCM. In order to purchase insurance informed by a genetic test result and nothing else, events would have to follow a very particular, and perhaps unlikely, order. We have, in effect, assumed that this is always the case, which is extreme. Furthermore, our approach would still be regarded as reasonable, if all patients received and followed the advice given by the genetic counsellors referred to in Lane et al. (2015) as follows: "In our study, counselors commonly recommended that patients secure all insurance needs before undergoing genetic testing" but also that "Generally, counselors reported advising patients that family history of diseases may have a greater impact on their insurability than genetic test results".

(e) **Allowance for selection bias**. Our study, or similar studies, relies on the genetic epidemiological literature, which evolves. Earlier studies revealing the

substrate of a genetic disorder were based on small numbers of affected individuals or families and high mortality rates were observed. Over time, we obtained the results of larger samples of less affected individuals, but who were still selected, which resulted in estimated mortality rates still overestimated as a result of selection bias. Carrying out population-based studies associated with genetic disorders is difficult since these disorders are relatively rare in the general population. Such studies based on genotype are still rarer, but evolving, compared to those based on the associated phenotype. There is no doubt that HCM is a good case study illustrating the evolving understanding of genetic disorders. For example, in Section 2.4.4, we showed that the 3–6% estimated HCM-related annual mortality rates between 1958–1990 dropped to the widely-cited annual mortality rate of about 1% when the manner of selection bias was significantly reduced. In the most recent studies, see also Section 3.9.7, the HCM-related annual mortality was even estimated to be about half of 1%. We also noted that, in Section 2.5.2, only four of twenty-two tested persons who carried a known HCM-related mutation had clinical HCM in Bick et al. (2012). They said in their paper: "This might reflect a lower sensitivity of population screening echocardiograms or *a lower disease penetrance in the general population compared to previously studied HCM families*" (emphasis added). Earlier actuarial studies conducted on genetics and insurance discussed the bias, namely *ascertainment bias*, resulting from the results of studies that only studied families and individuals which cause to the attention of epidemiologists. Lemaire et al. (2000) said (emphasis added):

"The results of our research should be applied with caution. They are based on the most recent data available from the medical literature, but new medical articles are published regularly that often provide very different estimates of BC and OC risks, depending on the demographic group studied. For instance, estimates of the lifetime probability of developing ovarian cancer for a woman with a BRCA1 mutation range from 11% to 84%. Also, *there might be a systematic bias in medical studies due to the selection of the sample, usually families with a strong family history* ".

Moreover, Macdonald et al. (2003*b*), a study of the impact of breast and ovarian cancer on critical illness insurance suggested reducing estimated onset rates by as much as 50% to 75% with regard to this bias. Note that we have not done so in our model. In Section 9.7.4, in the sensitivity analysis of fatal HCM, we present examples of selection bias in HCM in terms of the hazard rate of fatal HCM. In respect of all the sources of potential bias in the epidemiological studies, we are inclined to regard every value for premium increases we have calculated as likely an overstatement.

As a result, in Table 9.10, we present several examples of the impact of selection bias on the hazard rate of fatal HCM.

(f) **Improved treatment of HCM**. Even though we do not model the onset of HCM as a state in the adverse selection model, some might argue that the treatment of HCM will continue to improve and that will be another factor reducing mortality. Even if we ignore future developments in the treatment techniques, better diagnosis of clinical HCM would bring treatment to more individuals with HCM, so that will reduce mortality. If this were the case, the mean premium increases in Table 9.4 might be overstated.

9.7.3 Alternative Scenarios: $sc.17 - sc.20$

Based on our analysis of the significant factors (the model parameters) for adverse selection costs in Sections 9.7.1 and 9.7.2 above, we consider the combinations of these parameters in this section.

(a) In Table 9.7,

  (i) label $sc.17$ (a smaller market with mutation prevalence of 0.9%, and 'adverse selectors' taking out the average sum insured), and

  (ii) label $sc.18$ (a large market with mutation prevalence of 0.9%, and 'adverse selectors' taking ten times the average sum insured)

present the results of the combination of higher mutation prevalence (mutation prevalence being larger than the prevalence of clinical HCM, see point (b)

(i) in Section 9.7.2) in a smaller market where adverse selectors purchase the normal sum assured of £1 and a large market where adverse selectors purchase 10 times the normal sum assured of £1, see points (c) and (d) in Section 9.7.1 respectively.

A comparison of the results in labels *sc.*17 and *sc.*18 with those in labels *sc.*9 and *sc.*12 (the same assumptions in labels *sc.*17 and *sc.*18 but with mutation prevalence of 0.2%) in Table 9.5, respectively, shows that the prevalence of 'silent' HCM mutations diminish the adverse selection costs by a factor of about three. This conclusion overlaps with the conclusion of Section 9.3.1.

(b) Label *sc.*19 in Table 9.7 presents the results of the combination of a smaller market and high sums assured with mutation prevalence of 0.2%, see points (c) and (d) in Section 9.7.1 respectively. This scenario only focused on the impact of two combined significant pricing parameters, where adverse selectors purchase £10 sum assured in a smaller market, on the adverse selection costs, and found that increases in the adverse selection costs are somewhat high relative to those observed in the other scenarios.

Additionally, we present the results of a scenario, see label *sc.*20 in Table 9.8, consisting of a combination of the extreme values of the significant factors raising the adverse selection costs to considerable increases, namely the extreme case scenario. This scenario included assumptions of a smaller market, with mutation prevalence of 0.2%, extensive cascade genetic testing ($\lambda = 7.0$), 'severe' adverse selection and 'adverse selectors' taking ten times the average sum insured. This helps us to understand an approximate upper limit of adverse selection costs under extreme circumstances.

9.7.4 Sensitivity Analysis of Fatal HCM: *sc.*21 − *sc.*28

We present a sensitivity analysis of fatal HCM as follows:

(a) We first consider the magnitude of this epidemiological parameter in a smaller market, see label *sc.*21 in Table 9.9, where the annual hazard rate of fatal and

Table 9.7: Percentage increases in premium rates associated with the alternative scenarios. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.17	**0.9%**	**0.22**	$\mathbf{0.22}\boldsymbol{F^{b**}_x}$	0.6931472	$\rho^{nf\ddagger}_x$	0.0055	1.8	**1%**;£1	**1%**; £1	None	%0.0000	$(-0.0665, 0.0667)$	%0.0000	$(-0.0668, 0.0667)$
								1%;£1	**2%**; £1	Mild	0.0161	$(-0.0510, 0.0826)$	0.0052	$(-0.0618, 0.0720)$
								1%;£1	**25%**; £1	Severe	0.1167	(0.0406,0.1848)	0.0382	$(-0.0374, 0.1057)$
sc.18	**0.9%**	**0.22**	$\mathbf{0.22}\boldsymbol{F^{b}_x}$	0.6931472	ρ^{nf}_x	0.0055	1.8	5%; £1	5%; **£10**	None§	0.1843	(0.0952,0.2797)	0.0598	$(-0.0283, 0.1567)$
								5%; £1	10%; **£10**	Mild	0.2730	(0.1654,0.3668)	0.0886	$(-0.0186, 0.1805)$
								5%; £1	25%; **£10**	Severe	0.3698	(0.2472,0.4870)	0.1207	$(-0.0024, 0.2374)$
sc.19	0.2%	1	F^b_x	0.6931472	ρ^{nf}_x	0.0055	1.8	**1%**; £1	**1%**; **£10**	None§	0.4616	(0.2920,0.6290)	0.1558	$(-0.0066, 0.3184)$
								1%; £1	**2%**; **£10**	Mild	0.8875	(0.6619,1.0969)	0.3017	(0.0815,0.5212)
								1%; £1	**25%**; **£10**	Severe	3.5935	(3.2145,4.0096)	1.2391	(0.8450,1.6515)

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.
§ Adverse selectors purchase at normal rate, but higher sums assured.

Table 9.8: Percentage increases in premium rates associated with the extreme case scenario. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	$\zeta^n/\ \zeta^0$	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.20	0.2%	1	F^{b**}_x	0.6931472	$\rho^{nf\ddagger}_x$	0.0055	**7.0**	**1%**;£1	**25%**; **£10**	Severe	%8.3414	(7.8856,8.7829)	%2.8159	(2.3912,3.2220)

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.
‡ ρ^{nf}_x: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.

non-fatal HCM is 0.01 and zero respectively. We presented the results of the same assumptions in a large market shown in label *sc.*4 in Table 9.3.

A comparison of the scenario labelled by *sc.*21 with label *sc.*9 (a smaller market with the annual hazard rate of fatal and non-fatal HCM is 0.0055 and Table 3.6 respectively) in Table 9.5 shows that the impact of this adjustment on the mean premium increases is similar to that of the large market.

(b) In the scenario labelled by *sc.*4 in Table 9.3 (Section 9.3.4), we increased the annual hazard rate of fatal HCM at all ages from 0.0055 to 0.01, but we decreased the annual hazard rate of non-fatal HCM at all ages (Table 3.6) to zero. We observed that the adverse selection costs increase slightly compared to those in the baseline scenario, *sc.*0, see Table 9.1. However, a better comparison of *sc.*4 is the combination of the values of 0.0055 and zero for the annual hazard rate of fatal and non-fatal HCM, respectively.

The results are shown in label *sc.*22 in Table 9.9. The mean premium increases in *sc.*4 are 2.5 times of those in *sc.*22 under severe selection, even though they are still very small in both scenarios (*sc.*4 and *sc.*22). It seems that modelling non-fatal HCM stands as a significant factor for the adverse selection costs.

(c) We give examples of the adverse selection costs under the different circumstances of selection bias of the hazard rate of fatal HCM (see point (e) in Section 9.7.2)

The results are shown, when the insurer calculates premium rates assuming that the annual hazard rate of fatal and non-fatal HCM are 0.01 and zero respectively, when in reality two kinds considered:

(i) the annual hazard rate of fatal and non-fatal HCM are 0.0055 and Table 3.6 in labels *sc.*23, *sc.*25, and *sc.*27 in Table 9.10;

(ii) the annual hazard rate of fatal and non-fatal HCM are 0.0055 and zero in labels *sc.*24, *sc.*26, and *sc.*28 in Table 9.10.

The results are striking, especially when family history is allowed, because insurers might profit under adverse selection.

Table 9.9: Percentage increases in premium rates in respect to a sensitivity analysis of fatal HCM associated with that the annual hazard rate of non-fatal HCM is zero at all ages in the baseline scenario, *sc*.0, see Table 9.1. *sc*: Scenario. $p^{\mathcal{H}}$: Prevalence of HCM mutations at age 20 in the general population. $F^{eo}(20)$: Penetrance of early-onset HCM at age 20. $F^{lo}(x)$: Penetrance of late-onset HCM at ages 20–70. $\mu^{i02}_{x,z}$: Hazard rate of uptake of genetic testing at-risk relatives after an appearing proband in the family. ρ^{ijg}_x: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ^{ij8}_x: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. λ: Average number of children per family. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological Parameters							P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
sc	$p^{\mathcal{H}*}$	$F^{eo}(20)$	$F^{lo}(x)$	$\mu^{i02}_{x,z(0<z\leq 1)}$	ρ^{ijg}_x	ρ^{ij8}_x/μ^{ij8}_x	λ	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.21	0.2%	1	F^{b**}_x	0.6931472	**0**	**0.01**	1.8	**1%**;£1	**1%**; £1	None	%0.0000	$(-0.0442, 0.0487)$	%0.0000	$(-0.0435, 0.0489)$
								1%;£1	**2%**; £1	Mild	0.0532	(0.0078,0.1023)	0.0172	$(-0.0267, 0.0662)$
								1%;£1	**25%**; £1	Severe	0.3951	(0.3307,0.4616)	0.1252	(0.0660,0.1882)
sc.22	0.2%	1	F^b_x	0.6931472	**0**	**0.0055**	1.8	5%;£1	5%;£1	None	0.0000	$(-0.0236, 0.0254)$	0.0000	$(-0.0242, 0.0253)$
								5%;£1	10%;£1	Mild	0.0123	$(-0.0123, 0.0376)$	0.0040	$(-0.0206, 0.0289)$
								5%;£1	25%;£1	Severe	0.0257	$(0.0007, 0.0516)$	0.0082	$(-0.0158, 0.0345)$

* $p^{\mathcal{H}}$ uses baseline sub-population prevalences in the HCM population in Table 3.2.
** F^b_x: Baseline penetrance of late-onset HCM in Figure 3.4.

Table 9.10: Percentage increases in premium rates in respect to a sensitivity analysis of selection bias on the hazard rate of fatal HCM (point (e) Section 9.7.2). *sc*: Scenario. ρ_x^{ijg}: Hazard rate of non-fatal HCM per annum at all ages, the same for $g = 4, 5$, or 6. ρ_x^{ij8}: Hazard rate of fatal HCM per annum at all ages, the same for $j = 0, 1, 2$, or 3. μ_x^{ij8}: Hazard rate of fatal HCM per annum at all ages, the same for $j = 4, 5$, or 6. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	Premium Calculations			Actual Experience							
	Non-Fatal	Fatal	P.Rate;P.Sum	Non-Fatal	Fatal	P.Rate;P.Sum		Adverse	%Mean Premium Increases (QI,95%)			
	ρ_x^{ijg}	ρ_x^{ij8}/μ_x^{ij8}	$\zeta^n/\zeta^0/\zeta^{50}/\zeta^{100}$	ρ_x^{ijg}	ρ_x^{ij8}/μ_x^{ij8}	$\zeta^n/\ \zeta^0$	ζ^{50}/ζ^{100}	Selection	Family History: ✗		Family History: ✓	
sc.23	**0**	**0.01**	5%;£1	$\boldsymbol{\rho_x^{nf}}$‡	**0.0055**	5%;£1	5%;£1	None	%−0.2132	(−0.2369, −0.1887)	%−0.2357	(−0.2616, −0.2121)
	0	**0.01**	5%;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	5%;£1	10%;£1	Mild	−0.1873	(−0.2106, −0.1629)	−0.2404	(−0.2646, −0.2144)
	0	**0.01**	5%;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	5%;£1	25%;£1	Severe	−0.1583	(−0.1838, −0.1347)	−0.2460	(−0.2726, −0.2210)
sc.24	**0**	**0.01**	5%;£1	**0**	**0.0055**	5%;£1	5%;£1	None	−0.1887	(−0.2122, −0.1634)	−0.1747	(−0.1988, −0.1497)
	0	**0.01**	5%;£1	**0**	**0.0055**	5%;£1	10%;£1	Mild	−0.1764	(−0.201, −0.1512)	−0.1763	(−0.2001, −0.1516)
	0	**0.01**	5%;£1	**0**	**0.0055**	5%;£1	25%;£1	Severe	−0.1631	(−0.188, −0.1372)	−0.1787	(−0.2024, −0.1526)
sc.25	**0**	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**1%**;£1	None	−0.1911	(−0.2322, −0.1502)	−0.2107	(−0.2517, −0.1700)
	0	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**2%**;£1	Mild	−0.1481	(−0.1916, −0.1076)	−0.2197	(−0.2624, −0.1770)
	0	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**25%**;£1	Severe	0.1306	(0.0718,0.1889)	−0.2739	(−0.3330, −0.2160)
sc.26	**0**	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**1%**;£1	None	−0.1670	(−0.2051, −0.1236)	−0.1538	(−0.1919, −0.1120)
	0	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**2%**;£1	Mild	−0.1455	(−0.1836, −0.0998)	−0.1571	(−0.1956, −0.1105)
	0	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**25%**;£1	Severe	−0.0078	(−0.0552, 0.0428)	−0.1800	(−0.2266, −0.1292)
sc.27	**0**	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**1%**;**£10**	None§	0.2697	(0.1003,0.4367)	−0.3114	(−0.4724, −0.1511)
	0	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**2%**;**£10**	Mild	0.6947	(0.4695,0.9037)	−0.3984	(−0.6080, −0.1732)
	0	**0.01**	**1%**;£1	$\boldsymbol{\rho_x^{nf}}$	**0.0055**	**1%**;£1	**25%**;**£10**	Severe	3.3937	(3.0155,3.8090)	−0.8991	(−1.3240, −0.4820)
sc.28	**0**	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**1%**;**£10**	None§	0.0687	(−0.0406, 0.1835)	−0.1874	(−0.2982, −0.0701)
	0	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**2%**;**£10**	Mild	0.2829	(0.1328,0.4418)	−0.2196	(−0.3692, −0.0581)
	0	**0.01**	**1%**;£1	**0**	**0.0055**	**1%**;£1	**25%**;**£10**	Severe	1.6374	(1.3197,1.9453)	−0.4398	(−0.7347, −0.1380)

‡ ρ_x^{nf}: Baseline hazard rate of non-fatal HCM per annum at all ages in Table 3.6.
§ Adverse selectors purchase at normal rate.

Chapter 10

Comparison with Howard (2014)

10.1 Introduction

Howard (2014) models 13 genetic disorders (including HCM) for the Canadian life insurance market in which genetic test results are assumed not to be disclosed to life insurers. The model estimates the benefit claim costs and changes in the overall mortality experience because of adverse selection in the Canadian life insurance industry. This chapter discusses the main aspects of the model in Howard (2014) such as the model parameters, assumptions, and insurance costs in respect of HCM, in comparison with our model. Section 10.2 describes the baseline assumptions in respect of HCM in Howard (2014). Section 10.3 exhibits the actuarial models calculating the benefit claim costs and overall mortality experience under adverse selection in Howard (2014). Section 10.4 compares Howard (2014) with our study. Section 10.5 models lapses which appeared in Howard (2014), but have not considered in our model, so far.

10.2 Howard (2014): Baseline Assumptions

10.2.1 Prevalence, Penetrance and Mortality

The prevalence of HCM-related mutations is assumed to be 0.2% in the general population. (The size of the Canadian population is assumed to be 35 million.)

The penetrance of a genetic disorder is identified by the probability of a mutation carrier developing the associated disorder and being exposed to its mortality risk. This probability for HCM-related mutation carriers is assumed to be 69%.

Moreover, the mortality is classified in respect of the underwriting terms: 'sub-standard' (the mortality rate after the phenotype has developed) and 'standard' (the mortality rate in the absence of the phenotype). In other words, the individuals in the sub-standard group have higher mortality rates than those in the standard group. The annual mortality rate of HCM is assumed to be $q_x = 0.01$, see Sections 1.3 and 3.9.8, in all years after phenotype has developed. This is treated as a flat addition to the mortality rates of the standard group, the CIA 97-04 mortality tables.

Note that except for inherited breast cancer, the risks associated with the genetic disorders are assumed to be not gender-related.

10.2.2 Genetic Testing and Insurance Purchase

Since insurers are assumed not to be allowed to use genetic test results, it is assumed that 75% of the individuals who tested positive for any of the genetic disorders included in the model would attempt to purchase life insurance at $1 million sum assured (a 'normal' sum assured of $100,000 is assumed to be already in force, therefore positively-tested individuals are assumed to purchase an additional $900,000 sum assured). However, the model had more parameters associated with genetic testing such as the following.

(a) 'Predicted' is a model parameter referring to the proportion of positively tested individuals assumed to be detected in the underwriting process based on family history or early manifestation of the disorders at the time of insurance purchase. This proportion is 50% for HCM.

(b) 'Tested' is a model parameter referring to the proportion of the mutation carriers taking up genetic testing at an assumed average age. All the individuals supposed to be carrying the mutations causing these disorders are assumed to take up genetic testing at an average age specific to each genetic disorder in the model. This average age is 25 for HCM.

10.2.3 Lapse Rate

Except for Long QT syndrome, the lapse rate for the sub-standard and standard group is assumed to be 0.5% and 3% per annum in all years, respectively. See Section 10.5.2.

10.2.4 Interest Rate

The interest rate is assumed to be 4% per annum in all years.

10.2.5 Other

Other baseline assumptions such as the declination rate of insurance applicants, representative policy, conversion and expenses will not be significant for our purposes.

10.3 Howard (2014): Models and Results

10.3.1 The Cost Sub-model

The cost sub-model calculates the expected present value of the additional benefit claim costs under adverse selection. It compares the estimated adverse selection costs with $3.5 billion, the total amount of individual death claims in the Canadian life insurance industry in 2012.

(a) Under baseline assumptions, the expected present value of adverse selection costs is about 12% (or exactly 11.6%) ($405,455,952) of the total claim costs arising from all life insurance products in Canada ($3.5 billion). Also about 22% ($89,187,658) of the 12% is explained by HCM, which was the second highest single contribution to the costs. For comparison, only 1.3% ($5,363,834) of the 12% is explained by inherited breast cancer.

(b) Moreover, the expected present value of adverse selection costs as a percentage of the total claim costs would be:

    (i) 9.6% if the interest rate is 5% (instead of 4%);

(ii) 8.7% if 75% of the mutation carriers are tested (instead of all);

(iii) 11.7% if the standard group lapse rate is 0.5% (instead of 3%).

(iv) 12.9% if the sub-standard group lapse rate is 0% (instead of 0.5%); and,

(v) 12.9% if $1M additional sum assured is purchased (instead of $900K).

Note that these results were not shown separately for each genetic disorder as was done in the baseline scenario.

(c) Changes in the other model parameters such as expenses, declination rate, or premium rates give similar adverse selection costs in the baseline scenario, but mostly in a lower degree.

10.3.2 The Experience Sub-model

The experience sub-model calculates the expected increases in mortality rates with regard to the CIA mortality tables in which insurers classify mistakenly individuals in the sub-standard group as being in the standard group under adverse selection.

Under baseline assumptions, the overall mortality experience could increase by 43.8% (36% for males and 58% for females at ages 20–60). This figure would:

(a) decrease to the lowest level of 33.3% with the same change in the proportion of the mutation carriers being tested in point (b) (ii) in Section 10.3.1.

(b) increase to the highest level of 48.4% with the same change in the sum assured in point (b) (v) in Section 10.3.1.

10.3.3 An Alternative Scenario

In an alternative scenario, Howard (2014) estimated the adverse selection costs in which genetic test results are not disclosed to insurers up to $100,000 sum assured. Also, 25% of tested positive individuals are assumed to have already purchased an insurance sum assured of more than $100,000 or have not purchased any at all. It is observed that the additional claim costs are about 1.8% of the total claim costs and the overall mortality experience could increase by 3% for males and 8% for females.

Table 10.1: Percentage increases in premium rates associated with a change in the force of interest in the baseline scenario, *sc*.0, in our model, see Table 9.1. *sc*: Scenario.

	Force of Interest	Adverse Selection	%Mean Premium Increases (QI,95%) Family History: ✗		Family History: ✓	
sc.0	0.05	None	%0.0000	$(-0.0239, 0.0245)$	%0.0000	$(-0.0250, 0.0242)$
		Mild	0.0259	$(0.0025, 0.0503)$	0.0094	$(-0.0145, 0.0349)$
		Severe	0.0550	$(0.0294, 0.0786)$	0.0198	$(-0.0071, 0.0436)$
sc.29	0.04	None	0.0000	$(-0.0218, 0.0230)$	0.0000	$(-0.0227, 0.0228)$
		Mild	0.0232	(0.0017,0.0460)	0.0084	$(-0.0131, 0.0317)$
		Severe	0.0482	(0.0255,0.0699)	0.0174	$(-0.0062, 0.0390)$

These results are much less compared to those in the baseline scenario in Howard (2014).

The author applies this scenario is presumably because the Canadian law (GNDA) (Section 1.1) prohibiting insurers to use genetic test results; which were not specifically for insurers, and did not determine a threshold sum assured at which genetic test results are not disclosed to insurers.

10.4 Comparison

Our baseline adverse selection costs are very small (Section 9.2). Since there is no data available to parametrise the behavioural parameters in terms of purchasing insurance, the parametrisation of the majority of these parameters was conjectural. On the other hand, the parametrisation of the epidemiological (or biological) parameters may, even based on the epidemiological studies of HCM, lie within some reasonable range. Varying the model parameters may increase or decrease the adverse selection costs. This comparison seeks out factors that may increase the costs.

This comparison is needed due to the disparity between Macdonald & Yu (2011) and Howard (2014) in terms of the reported adverse selection costs (Section 1.1). The largest part of the costs in Howard (2014) were explained by cardiomyopathies in which they were not modelled in Macdonald & Yu (2011). Direct comparison of both models is not possible; however, it would be useful to compare the percentage premium increases caused by HCM in Howard (2014) with those in our study.

(a) Point (a) in Section 10.3.1 suggests that increased premiums of about 2.5% caused by HCM in Howard (2014) is roughly equivalent to our model.

(b) Section 10.2.1 presents the assumptions for the epidemiological (or biological) parameters in Howard (2014). Our baseline assumptions do not significantly differ from these except that the annual mortality rate which we assume is about the half of $q_x = 0.01$ Howard (2014) assumed, see Sections 1.3, 3.9.8, and 10.2.1. Fortunately, it turned out the adverse selection costs by this difference are not substantial (Section 9.3.4). Although this section does not seek out factors that diminish adverse selection costs, bear in mind that modelling the prevalence of 'silent' mutations (Section 9.3.1) reduced the adverse selection costs considerably, by a factor of about three in our study.

(c) Section 10.2.2 presents the assumptions for the pricing parameters (or behavioural parameters) associated with the individuals with positive genetic test results in Howard (2014). In our study (see Chapter 9), we assumed 5% and 1% for 'annual hazard rate of normal purchase' in a large insurance market and a small market, respectively. In both cases, under severe adverse selection, we increase the normal purchase rate per annum to 25%, as Macdonald & Yu (2011) noted that "this assumption is deliberately high; it implies that about 91% of at-risk people would buy insurance in 10 years in both large and small markets". This might be a good comprasion with Howard (2014)'s assumption of the purchase rate of the adverse selectors shown in Section 10.2.2; moreover, Howard (2014) also assumed that adverse selectors would purchase insurance at \$1,000,000 sum assured, larger by a factor ten than the \$100,000 normal sum assured. Label *sc.*12 in Table 9.5 for a large market and label *sc.*19 in Table 9.7 for a smaller market show the results in our model (with mutation prevalence of 0.2%) in which adverse selectors purchase insurance at £10 sum assured, larger by a factor of ten than £1 normal sum assured.

(d) Howard (2014), see Section 10.2.3, modelled lapses that we did not model so far. However, having a smaller insurance purchase rate in our model, see

Section 9.5.1, would achieve a similar effect to the combination of a higher insurance purchase rate with a notable lapse rate. Thus, the results in label *sc*.9 in Table 9.5 gives a fairer comparison with Howard (2014) rather than the results in the baseline scenario labelled by *sc*.0 in Table 9.1. However, we will model lapses (Section 10.5) in respect of the reported lapse rates in Howard (2014) (Section 10.2.3), for a closer comparison.

(e) Howard (2014), see Section 10.2.4, assumed the interest rate to be 4% per annum in all years. We replace the 0.05 baseline force of interest per annum by 0.04 in the baseline scenario, *sc*.0, see Table 9.1 in our model. The results are shown in Table 10.1. The adverse selection costs do not significantly change.

(f) In Sections 10.3.1 and 10.3.2, we reported the sensitivity analysis for the benefit claim costs and the overall mortality experience, respectively, undertaken in Howard (2014). In Section 10.3.3, we presented an alternative scenario in Howard (2014) in which insurers are assumed to be allowed to have genetic test results for sums assured higher than $100,000.

None of the results, either in the sensitivity analysis or the alternative scenario in Howard (2014) was reported separately for HCM. However, the sensitivity tests did not change adverse selection costs by much, but the alternative $100,000 'ceiling' scenario did.

10.4.1 Highlights

In this section, we highlight some important points in this comparison. Our results include a range of scenarios (Chapter 9) compared to Howard (2014)'s one scenario (baseline), subject to some rather minor sensitivity analysis.

(a) Howard (2014) does not consider at all the status and current stage of evolution of the HCM literature as we did in this study.

(b) Howard (2014) considers only extreme sums assured such as adverse selectors purchase insurance at 10 times of the normal sum assured. In our study, 2 and 4 times of the normal sum assured for adverse selectors were considered.

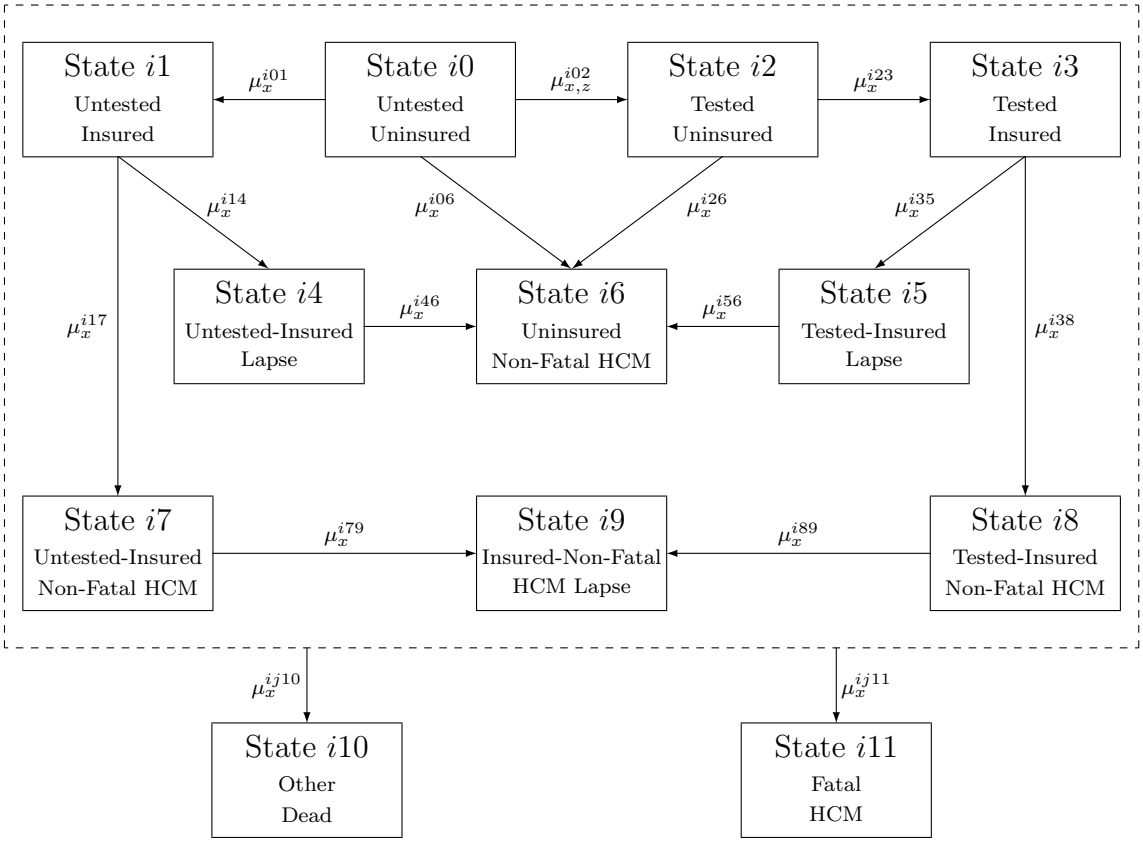


Figure 10.1: A mathematical model, including lapse, of adverse selection in HCM of a person in the ith of several sub-populations defined by HCM genotype in a life insurance market. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family.

10.5 Modelling Lapses in Hypertrophic Cardiomyopathy (HCM) for Life Insurance

10.5.1 The Lapse Model

In this section, we model lapses. Lapsing means that a policyholder ceases the insurance policy by not continuing to pay premiums due for the rest of the policy term, and abandons in turn the benefits of the policy.

In Figure 10.1, we present a mathematical model, including lapse state(s), of adverse selection in HCM in a life insurance market, called the lapse model.

The lapse model is similar to the adverse selection model (Chapter 7) in that it keeps all the sub-populations in the adverse selection model together with their subdivisions into the HCM and non-HCM population and families, see Figures 7.2

and 7.3, respectively. On the other hand, the state space is increased in each sub-population by adding the new model states, namely lapse state(s) (note that the state space is same for each sub-population).

The methodology of the lapse model in terms of parametrising the model by the transition intensities; computing the occupancy probabilities from these intensities; and calculating actuarial losses in the model was largely described in Chapters 3, 5, 7, and 8 so we do not go into detail here, except the following new features of the lapse model.

(a) Note that the transitions out of the dashed box in Figure 10.1 represent the fact that an individual can move into a death state from any model state at any time.

(b) Note also that lapses can be modelled in several ways. For example, an insurer might agree to pay back some sum assured or a cash value to the policyholder as a return of the policyholder's contribution since the inception of the policy (Dickson et al. 2013). Also, a policyholder might re-purchase insurance after some time has elapsed (Subramanian et al. 1999). These are not considered in the lapse model.

(c) Here, instead, we model lapses, as we modelled changing purchase behaviour of the individuals in HCM families, after family history appears (there is a proband in the family) or a genetic test result known to the individuals (Chapter 7). Doing so is consistent with the aim of our study.

Therefore, we adopt the information classes defined in the adverse selection model, see Section 7.5.1. See also Tables 7.3 and 7.4 in which we present the information classes for each family member in an HCM family based on the events in the family. The definition of the information classes in the lapse model is the same as for the adverse selection model.

For example, suppose a person is in an HCM family, with no family history (such as no proband in the family), neither tested nor yet having purchased insurance (or in the untested-uninsured state). Then the information class of

this person is ζ^n. After that the person purchases insurance with no family history having appeared nor having being tested, so this person is still in the information class ζ^n. Then the carrier parent of the person becomes a proband in the family. Then this person is now in information class ζ^{50}. And the lapse behaviour might change based on this. See Section 10.5.2 for the assumed behavioural parameters for lapse under different information classes. Note that a person in state $i7$, untested-insured non-fatal HCM is assumed to be in information class ζ^{100} since HCM clinically exists now and lapse behaviour might change based on that.

In Figure 10.2, we represent the life history of the individuals in risk sub-population $i = 0$. These individuals are always members of non-HCM families, as their rates of insurance purchase, and lapse always normal. The unshown transition intensities can be found in Figure 7.6.

The life history of an individual in an HCM family based on the lapse decisions of this individual is shown in Figures 10.3, 10.4, 10.5, and 10.6. The lapse transition intensities in these figures correspond to the information classes and the unshown transition intensities can be found, respectively, in Figures 7.7, 7.8, 7.9, and 7.10.

10.5.2 The Parametrisation

We follow the reported lapse rates in Howard (2014), consistent with the aim of this chapter. As we stated in Section 10.2.3, Howard (2014) categorises the lapse rates in two categories:

(a) The lapse rate of 'sub-standard' lives, representing the mutation carrier individuals exposed to higher mortality (after the phenotype has developed) might have a lower than normal lapse rate, depending on the information they have. In the lapse model, this lapse rate applies to insured individuals in information classes ζ^{50} and ζ^{100}.

(b) The lapse rate of 'standard' lives, representing the non-mutation carrier indi-

Table 10.2: Hazard rates of lapse per annum at ages 20–60, relying on the reported lapse rates in Howard (2014), depending on the information class of insured person in the lapse model. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Note that an untested person suffered non-fatal HCM after purchasing insurance is assumed to be in information class ζ^{100} since HCM clinically exists now and lapse behaviour might change based on that.

Adverse Selection	Information Class ζ^n	ζ^{50}	ζ^{100}	ζ^0
None	3%	3%	3%	3%
Baseline	3%	0.5%	0.5%	3%
Sensitivity	3%	0%	0%	3%

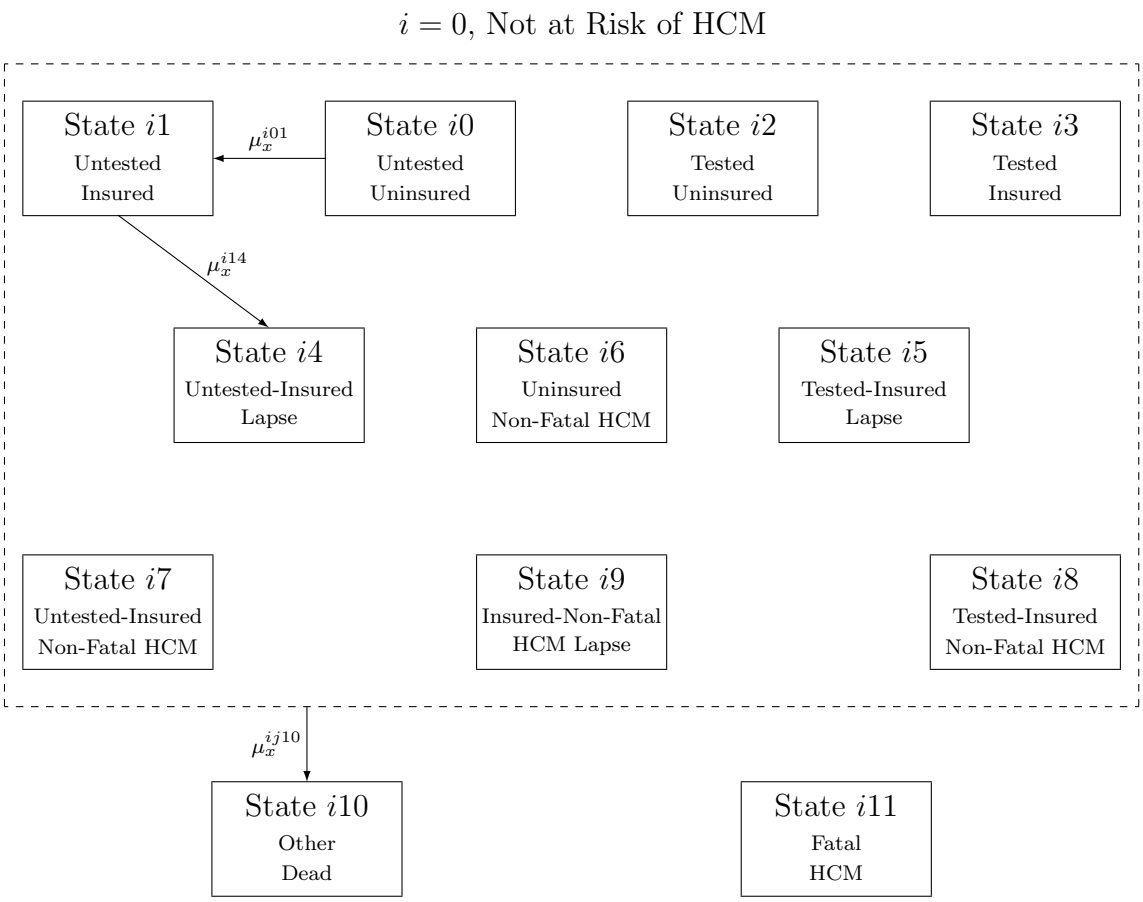
viduals only exposed to the population mortality rates, is always normal. In the lapse model, this lapse rate applies to insured individuals in information classes ζ^n and ζ^0.

Based on points (a) and (b) above (which are expressed in our terminology, not Howard (2014)'s), Howard (2014) assumes (as a baseline) lapse rates of 0.5% and 3% per annum in all years for the 'sub-standard' and 'standard' lives, respectively. Howard (2014) also runs a sensitivity analysis assuming a lapse rate of 0% for the 'sub-standard' lives, see point (iv) in Section 10.3.1. We will also consider the implication of this sensitivity analysis because it is conservative for our purposes. In Table 10.2, we present the lapse (hazard) rates depending on the information classes in the lapse model.

10.5.3 The Results and Comparison

We present the baseline assumptions for the lapse model in Table 10.3 in which we combine the baseline lapse (hazard) rates in Table 10.2 with the baseline assumptions in the adverse selection model reported in Table 8.1.

The results in respect of increased mean premium rates, see Chapter 9, are shown in Table 10.4. Modelling lapse seems to be a significant factor amplifying adverse selection costs, in particular the mean premium increases are more than 1% when



Fixed Rates:

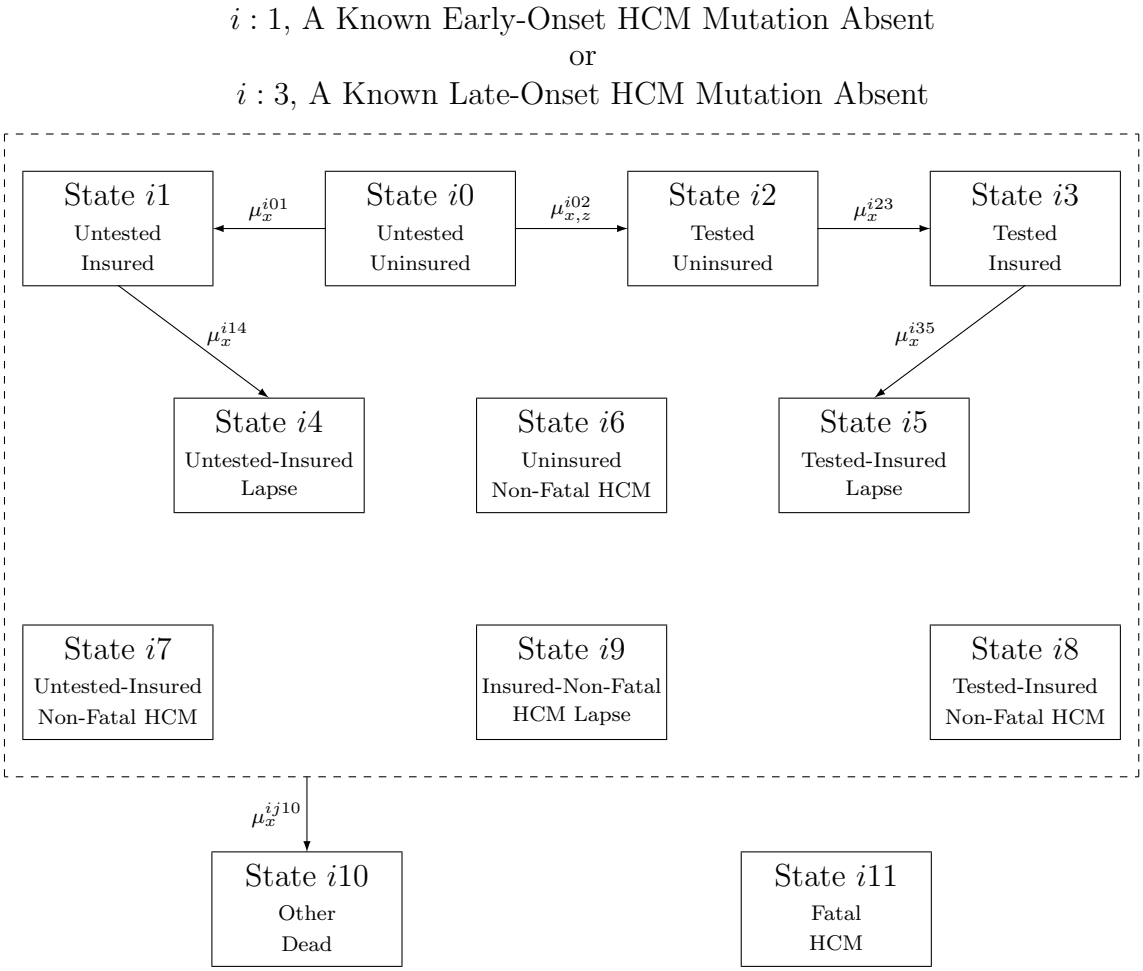
- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

Figure 10.2: A mathematical model of a life history of an individual who is a member of the $i = 0$ risk sub-population. See Figure 7.6 for all transitions.

family history is not known to insurers. This model parameter is worth adding to the list of factors amplifying adverse selection costs in Section 9.7.1.

Table 10.3: Baseline assumptions for the lapse model parameters.

Epidemiological Parameters	Table 3.10	Section 3.10
Prevalence of non-HCM mutations in the general population at age 20	0.998	Section 3.7.1
Prevalence of HCM mutations in the general population at age 20	0.002	Section 3.7.1
Prevalence of known early-onset mutations in the HCM population at birth	0.5	Section 3.7.2
Prevalence of known late-onset mutations in the HCM population at birth	0.1667	Section 3.7.2
Prevalence of unknown early-onset mutations in the HCM population at birth	0.25	Section 3.7.2
Prevalence of unknown late-onset mutations in the HCM population at birth	0.0833	Section 3.7.2
Penetrance of early-onset HCM at age 20	100%	Section 3.8.1
Penetrance of late-onset HCM at ages 20–70	Figure 3.4	Section 3.8.2
Hazard rate of fatal HCM per annum for all ages	0.0055	Section 3.9.7
Hazard rate of non-fatal HCM per annum for all ages	Table 3.6	Section 3.9.7
Hazard rate of all other death per annum for all ages	Figure 3.9	Section 3.10
Hazard rate of testing in one year at ages 0–70 if proband exists in family	0.6931472	Section 4.3.2
Average number of children per family (λ)	1.8	Section 5.6.1
Premium Rates	Equation (7.1)	Section 7.4.3
Benefit (Sum Assured)	£1	Section 7.4.5
Force of interest per annum	0.05	Section 7.4.5
Hazard rate of normal purchase per annum at ages 20–60	0.05	Section 7.5.1
Hazard rate of purchase per annum at ages 20–60 based on information class of individuals	Table 7.2	Section 7.5.1
Hazard rate of normal lapse per annum at ages 20–60	0.03	Section 10.5.2
Hazard rate of lapse per annum at ages 20–60 based on information class of individuals	Table 10.2	Section 10.5.2



If no proband exists in family:
$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

If carrier parent becomes proband with a known mutation:

(a) *m is a spouse of carrier parent, $\zeta_m = \zeta^n$;*
- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

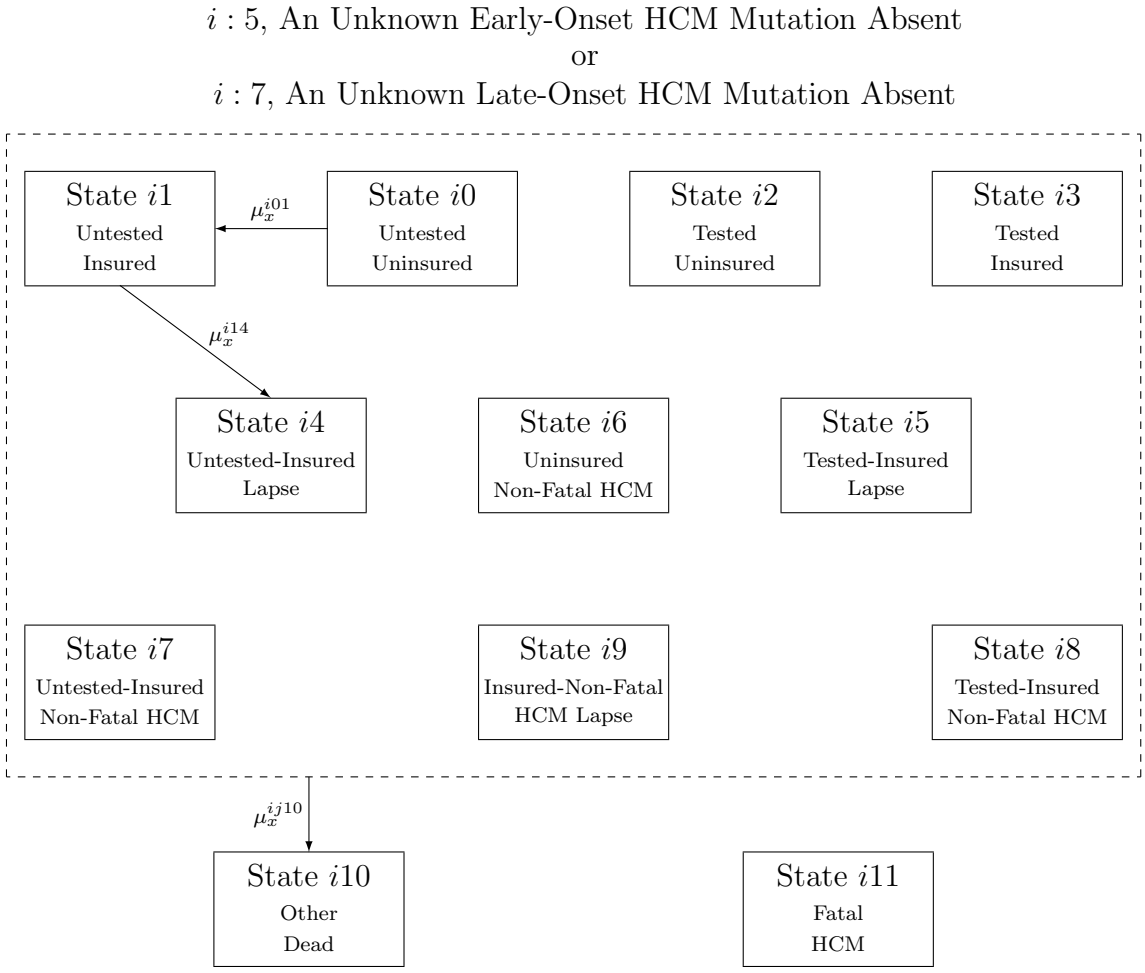
(b) *m is a non-carrier child of carrier parent, $\zeta_m = \zeta^{50}$; or, $\zeta_m = \zeta^0$.*
- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (might change) per annum at ages 20–60.
- $\boldsymbol{\mu_x^{i35}}$ = normal lapse rate (not change) per annum at ages 20–60.

If a carrier child becomes proband with a known mutation:

(a) *m is a spouse of carrier parent not tested nor become a subsequent proband; or, m is a non-carrier sibling of the carrier child, $\zeta_m = \zeta^{50}$; or, $\zeta_m = \zeta^0$.*
- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (might change) per annum at ages 20–60.
- $\boldsymbol{\mu_x^{i35}}$ = normal lapse rate (not change) per annum at ages 20–60.

(b) *m is a spouse of carrier parent tested or become a subsequent proband, $\zeta_m = \zeta^0$; or $\zeta_m = \zeta^n$;*
- $\boldsymbol{\mu_x^{i14}} = \boldsymbol{\mu_x^{i35}}$ = normal lapse rate (not change) per annum at ages 20–60.

Figure 10.3: A mathematical model of a life history of an individual m, a non-carrier member in which one parent carries a known HCM mutation, in the $i = 1$ or $i = 3$ risk sub-populations in the lapse model of HCM for life insurance. See Figure 7.7 for all transitions. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, 7.4, and 10.2 for information classes.



If no proband exists in family:

$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

If carrier parent becomes proband with a known mutation:

(a) *m is a spouse of carrier parent, $\zeta_m = \zeta^n$;*

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

(b) *m is a non-carrier child of carrier parent, $\zeta_m = \zeta^{50}$;*

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (might change) per annum at ages 20–60.

If a carrier child becomes proband with a known mutation:

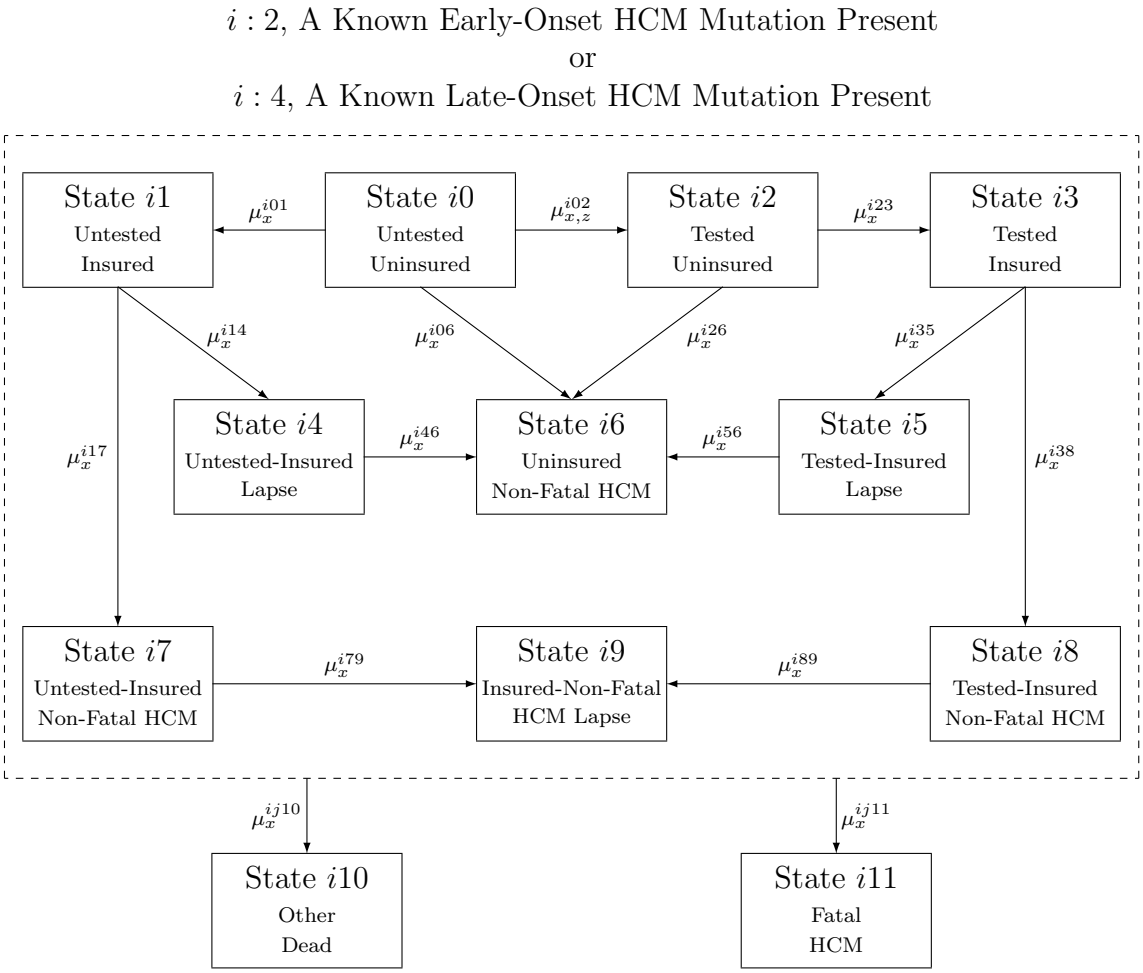
(a) *m is a spouse of carrier parent not become a subsequent proband; or, m is a non-carrier sibling of the carrier child, $\zeta_m = \zeta^{50}$;*

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (might change) per annum at ages 20–60.

(b) *m is a spouse of carrier parent become a subsequent proband, $\zeta_m = \zeta^n$;*

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

Figure 10.4: A mathematical model of a life history of an individual m, a non-carrier member in which one parent carries an unknown HCM mutation, in the $i = 5$ or $i = 7$ risk sub-populations in the lapse model of HCM for life insurance. See Figure 7.8 for all transitions. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, 7.4, and 10.2 for information classes.



If no proband exists in family:
$\zeta_m = \zeta^n$;

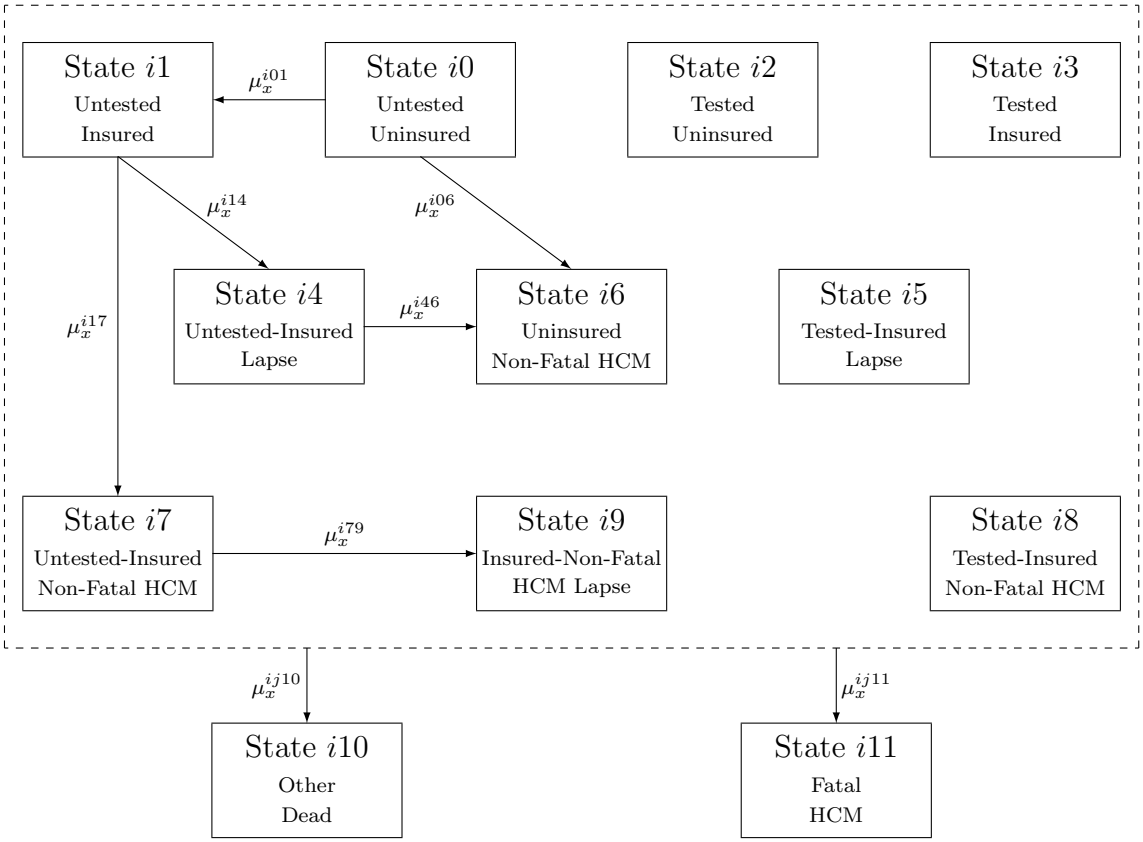
- $\boldsymbol{\mu_x^{i14}}$ = normal purchase rate (not change) per annum at ages 20–60.

If a proband exists (who not matter) with a known mutation in family:
$\zeta_m = \zeta^{50}$; or, $\zeta_m = \zeta^{100}$;

- $\boldsymbol{\mu_x^{i14} = \mu_x^{i79} = \mu_x^{i35} = \mu_x^{i89}}$ = normal lapse rate (might change) per annum at ages 20–60.

Figure 10.5: A mathematical model of a life history of an individual m, a carrier member in which one parent carries a known HCM mutation, in the $i = 2$ or $i = 4$ risk sub-populations in the lapse model of HCM for life insurance. See Figure 7.9 for all transitions. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband exists in the family. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, 7.4, and 10.2 for information classes.

$i : 6$, An Unknown Early-Onset HCM Mutation Present
or
$i : 8$, An Unknown Late-Onset HCM Mutation Present



If no proband exists in family:
$\zeta_m = \zeta^n$;

- $\boldsymbol{\mu_x^{i14}}$ = normal lapse rate (not change) per annum at ages 20–60.

If a proband exists (who not matter) with a known mutation in family:
$\zeta_m = \zeta^{50}$; or $\zeta_m = \zeta^{100}$ (because of state $i7$, untested-insured non-fatal HCM)

- $\boldsymbol{\mu_x^{i14} = \mu_x^{i79}}$ = normal lapse rate (might change) per annum at ages 20–60.

Figure 10.6: A mathematical model of a life history of an individual m, a carrier member in which one parent carries an unknown HCM mutation, in the $i = 6$ or $i = 8$ risk sub-populations in the lapse model of HCM for life insurance. See Figure 7.10 for all transitions. ζ_m refers to an information class for individual m. See Section 7.5.1 and Tables 7.2, 7.3, 7.4, and 10.2 for information classes.

Table 10.4: Percentage increases in premium rates associated with lapsing in our model. *sc*: Scenario. P.Rate: Hazard rate of purchase per annum at ages 20–60. P.Sum: Purchased sum assured. L.Rate: Hazard rate of lapse per annum at ages 20–60. ζ^n: No knowledge of any HCM risk in the family/a non-carrier parent with his/her spouse becoming proband. ζ^{50}: Believe themselves to be at 50% risk of carrying an HCM mutation. ζ^{100}: As a result of genetic testing, knows they carry an identical mutation. ζ^0: As a result of genetic testing, knows they do not carry an identical mutation. Force of interest is assumed to be 5% per annum at all ages.

	The Epidemiological	P.Rate;P.Sum		L.Rate		%Mean Premium Increases (CI,95%)			
sc	Parameters	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	ζ^n/ ζ^0	ζ^{50}/ζ^{100}	Family History: ✗		Family History: ✓	
sc.30	*sc*.0, Table 9.1	5%;£1	5%;£1	3%	3%	%0.0000	$(-0.0348, 0.0337)$	%0.0000	$(-0.0349, 0.0332)$
		5%;£1	5%;£1	3%	0.5%	0.0244	$(-0.0074, 0.0546)$	0.0125	$(-0.0194, 0.0421)$
		5%;£1	5%;£1	3%	0%	0.0306	$(-0.0021, 0.0605)$	0.0156	$(-0.0168, 0.0448)$
		5%;£1	25%;£1	3%	0%	0.1175	(0.0803,0.1575)	0.0476	(0.0088,0.0887)
sc.31	*sc*.12, Table 9.5	5%;£1	25%;£10	3%	3%	1.1861	(1.0021,1.3744)	0.4206	(0.2361,0.6136)
		5%;£1	25%;£10	3%	0%	1.6664	(1.4460,1.9081)	0.5911	(0.3789,0.8225)

Chapter 11

Discussion and Conclusions

11.1 The Epidemiology of HCM

We found that the prevalence of 'silent' mutations has a significant impact on adverse selection costs (Section 9.3.1). Therefore, if genetic testing is spread through the general population, it might reveal 'silent' mutations. For example, in HCM, the known mutation prevalence of 0.6% reported in Bick et al. (2012), which, extrapolated, may approach 1% is much higher than 0.2% the widely cited for the prevalence of clinical HCM. As a result, the penetrance is correspondingly reduced, see the quotation from Bick et al. (2012) in point (e) in Section 9.7.2. We might conclude the motivation to purchase insurance is less well-determined.

(a) This motivation could bring more individuals into the insurance pool who want to over-insure themselves (see point (c) in Section 9.3.1). But, in reality, the total number of the individuals affected by any HCM-related event will not change. Consequently, the cost of adverse selection is significantly reduced.

(b) This motivation, in the absence of additional information of a non-genetic nature, could also diminish any incentive to over-insure, in particular it might be a disincentive to stake a large quantity of money on insurance premiums in a gamble, the outcome of which looks less attractive.

Also, HCM differs from most the 'classical' genetic disorders in the actuarial literature (such as Huntington disease and inherited cancers—good references for these

disorders and their implications to critical illness and life insurance are Gutiérrez & Macdonald (2004) and Macdonald et al. (2003*b*)) in that the onset of the associated phenotype starts at early ages in life in a large proportion of cases, and is diagnosed by imaging machines (Section 2.2.2). If the disorder is diagnosed, it would be underwritten as a pre-existing condition. If insurance applicants want to use their genetic test results as a financial opportunity, they should have this genetic information in isolation without knowing of any other indication, which must be disclosed to an insurer, showing that HCM might be clinically-present. This might be another factor, which should be borne in mind, in diminishing the adverse selection costs arising when insurers are not allowed to use genetic test results.

The evolution of the epidemiology of HCM is still in progress. Apart from some studies of prevalence, all of it is based on selected populations. Actuarial studies should be aware of such biases. It is still too early to say where the epidemiology of HCM currently is. For example, non-fatal HCM events have been included as endpoints and may mistakenly be counted as 'actual' deaths, see Sections 2.4.2 and 2.4.4. Fortunately, our study showed that the effect of this appears not to be large, see Section 9.3.4. However, this bias might exist for other cardiomyopathies. Especially, Howard (2014) reported that the annual mortality rate of Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) (Section 1.2.3) (the most expensive disorder in his study) was 2.3%. If this bias was to exist for ARVC, its effect on the adverse selection costs, unlike HCM, might be significant. This is a subject for future research.

11.2 Range of Premium Increases

In Chapter 9 and Section 10.5.3, we presented the results with regard to the mean premium increases (to recoup adverse selection costs) under many adverse selection scenarios. These results showed adverse selection costs might lie in a large spectrum highly depending on the assumptions. For example,

(a) In one scenario labelled by *sc.*1 in Table 9.3, mean premium increases were 0.0029% in which assumptions were a large market, with family history al-

lowed, mutation prevalence of 0.9%, less extensive cascade genetic testing ($\lambda = 1.8$), 'mild' adverse selection, and 'adverse selectors' taking out the average sum insured. Note that assuming that individuals in information class ζ^{50} purchase less than those in information class ζ^{100} gives adverse selection costs even lower than this level (see label *sc.*15 in Table 9.6). Also, note that, we even observed 'negative' adverse selection costs (Table 9.10) when we take into consideration selection bias (point (e) in Section 9.7.2).

(b) In another scenario labelled by *sc.*20 in Table 9.8, in which assumptions representing much more extreme scenario were a smaller market, with family history not allowed, mutation prevalence of 0.2%, annual hazard rate of fatal HCM-related events at all ages of 0.0055, extensive cascade genetic testing ($\lambda = 7.0$), 'severe' adverse selection and 'adverse selectors' taking ten times the average sum insured, our mean premium increases were 8.3414%. This would appear to be far in excess of those suggested by Howard (2014). If family history is allowed, the mean premium increases decrease to 2.8159%.

This indicates the range of possibilities. Recall from Section 10.4 that premium increases of 2.5% or over in our model might be comparable to those in Howard (2014) even though exact comparison is not possible.

Note that we use a large number of assumptions to parametrise our model during the thesis. They are sometimes chosen in the absence of reliable studies. In such cases, we generally choose them 'conservatively', which would tend to lead an increase in adverse selection costs suffered by insurers.

11.3 Significant Factors

We already pointed out amplifying and diminishing factors in adverse selection costs in Section 9.7. Here we summarise them.

The following factors amplified adverse selection costs (Section 9.7.1):

- family history being unavailable in underwriting (Section 9.2);

- cascade genetic testing extending significantly beyond the nuclear family, approximated by an increased average family size (Section 9.4.2);
- a much smaller life insurance market (which may stand as a proxy for significant lapse rates in a larger market) (Section 9.5.1). See also our results with the lapse rates reported in Howard (2014) in Section 10.5.3; and
- 'adverse selectors' taking out extremely large sums insured (Section 9.5.2).

Otherwise, the following factors would tend to limit or reduce adverse selection costs (Section 9.7.2):

- family history being available in underwriting (Section 9.2); moreover, the importance of the use of family history has been realized by insurers. For example, "One of the main reasons the Association of British Insurers, from 1996 onwards, reached an agreement with the UK Government not to use genetic test results, was that it feared that a government-imposed ban on genetic tests might extend to the use of family history (personal communication with Professor Angus Macdonald)",
- the prevalence of 'silent' HCM mutations (Section 9.3.1);
- the individuals in information class ζ^{50} purchasing at a lower rate than those in information class ζ^{100} under adverse selection (Section 9.6) (note that we did not model different sums assured for the individuals in information classes ζ^{50} and ζ^{100} under adverse selection);
- including the onset of clinical HCM as an event in the model, instead of assuming $F(20) = 1$ for early-onset mutations (Section 9.3.1);
- unknown mutations being predominantly late-onset (Section 9.3.3.1);
- epidemiology based on unselected populations becoming available;
- elimination of selection and ascertainment biases from epidemiological studies; and

- improved treatments for HCM.

We do not model the last five of these explicitly, but they should be borne in mind. However, in Table 9.10, we present examples of the effect of selection bias (point (e) in Section 9.7.2) on adverse selection costs. Particularly, two factors become prominent in this discussion.

(a) **Larger sums insured**. It is obvious that adverse selection costs can be increased without limit, if 'adverse selectors' take out large enough sums insured. It is fair to ask where the motivation of adverse selectors comes from to view life insurance as a financial opportunity rather than an insurance need. It seems that their motivation would have to be quite different from the motivation to "... secure all insurance needs before undertaking genetic testing" (Lane et al. 2015). If we think about two annual hazard rates of HCM-related mortality used through the thesis: 0.01 and 0.0055 (Section 3.9.8), about 2/3 and 4/5, respectively, of individuals would survive for forty years. Thomas (2012) refers to "the fallacy of the one-shot gamble" in noting the lack of evidence for adverse selection of genetic origin in the UK, despite ten years (by now almost twenty years) of the opportunity existing. Such a gamble might be monetized if a large number of mutation carriers were organized and financed by some outside agency, to buy life insurance in large amounts. Though, any such scheme in reality might be a target of regulation, as having no place in a well-functioning life insurance market. Legal or not, the scheme's attraction would be proportionate to its promoters' belief in the most extreme assumptions.

(b) **Mutation prevalence**. We modelled the prevalence of 'silent' mutations, in other words, the large difference between mutation prevalence and the prevalence of clinical HCM (Section 9.3.1). The existing epidemiology is subject to even more selection and ascertainment biases (see point (e) in Section 9.7.2) that are unobservable and unknown, but are all in one direction. We could have taken into account some of these biases, by reducing onset rates although

by how much would be arbitrary. However, since adverse selection costs using any reasonable parameters look like small enough, we have not done so.

11.4 Conclusions

We modelled cascade genetic testing in HCM and this enabled us to explore possible adverse selection costs in respect of the percentage mean premium increases to recoup the costs.

(a) We found that the range of the mean premium increases to be very large. The very highest are comparable to, or may even be much in excess of, those suggested in Howard (2014). We also noted that, in Section 10.4.1, Howard (2014) presents basically the results of one scenario (baseline) with a minor sensitivity analysis. Compared to Howard (2014):

  (i) We obtained a broad range of results because we considered 'all the status and current stage of evolution of the HCM literature'.

  (ii) We also considered different sums assured for adverse selectors rather than only extreme sums assured (such as 10 times of the normal sum assured).

(b) We found that very high sums assured (such as 10 times of normal) cause significant increases in adverse selection costs (Section 9.5.2). We think this assumption is debatable. The question of whether it would be possible to monetize a 'one-shot gamble' (Thomas (2012), see point (a) in Section 11.3) would be an interesting topic for research.

(c) We found that some epidemiological features of HCM would reduce the possible adverse selection costs.

  (i) The major one is the prevalence of 'silent' mutations. Prevalence of HCM-related mutations is much higher than that of clinical HCM, suggesting lower penetrance in unselected populations (Section 9.3.1).

(ii) The other is the general practice among authors of HCM survival studies of using endpoints that include a substantial proportion of non-fatal events (Sections 2.4.2 and 2.4.4). Distinguishing them from fatal ones is probably less important for the epidemiologists, but is important for life insurance applications.

The impact of this bias on the adverse selection costs caused by HCM turned out to be not so high (Section 9.3.4). However, if this bias exists for other cardiomyopathies, such as ARVC (the most expensive disorder in Howard (2014)) (Sections 1.2.3 and 11.1), the significance of this bias for the possible adverse selection costs might be even more than in respect of HCM. This a subject for future research.

The impact of this bias on adverse selectors might be that they see purchasing high sums assured insurance as a less attractive financial opportunity (or they doubt the attractiveness of the 'one-shot gamble').

(iii) We have not reduced costs further by allowing for probable, but unquantified, biases in the epidemiological literature. These have evidently reduced over time but are unlikely to have disappeared.

HCM is regarded as the most prevalent of several dominantly-inherited cardiomyopathies and ion-channelopathies. Cascade genetic testing is the form of testing used in clinical practice. Our model should be capable of estimating insurance losses, in respect of each of these conditions, under adverse selection. This is a subject for future research.

Appendix A

Multiple-State Multiple-Population Models in Life Insurance

A.1 A Multiple-State Multiple-Population Model in Life Insurance

We extend the multiple-state model in Section 6.5 into a multiple-state multiple-population model analogous to our adverse selection model in Chapter 7. Therefore:

(a) Let a multiple-state multiple-population model consist of model states, which may include insured state(s), labelled by ij in which:

  (i) Label i denotes a sub-population, and

  (ii) Labels j, k, etc. denote sub-states in any sub-population i where transitions may be possible between states ij and ik, $j \neq k$. Note that each sub-population i contains a copy of the same state space.

(b) Assume transitions between different sub-populations are impossible.

(c) Suppose individual r starts at state $i0$ at calendar time zero with probability p_{i0}, so $\sum_i p_{i0} = 1$ (Section 3.3.1).

In what follows, we refer to the notations in Chapter 6.

- $N_r^{ijk}(t)$: The number of transitions of individual r from state ij to state ik up to and including calendar time t.

- $A_r^{ijk}(t)$: The lump sum payment (assumed to be previsible and possibly zero) payable to individual r under the transition $ij \to ik$ at calendar time t.

- $I_r^{ij}(t)$: An indicator function as follows:

$$I_r^{ij}(t) = \begin{cases} 1, \text{ individual } r \text{ is in state } ij \text{ at calendar time } t^-; \\ 0, \text{ otherwise.} \end{cases} \tag{A.1}$$

- $a_r^{ij}(t)$: The rate per annum of premium payment (assumed to be previsible and possibly zero) made if individual r is in state ij at calendar time t^-.

- $M_r^{ijk}(t)$: The counting process martingale of individual r in state ij as follows:

$$M_r^{ijk}(t) = N_r^{ijk}(t) - \int_0^t I_r^{ij}(w)\mu_{r,w}^{ijk} dw. \tag{A.2}$$

  where $\mu_{r,t}^{ijk}$ is the transition intensity between states ij and ik, $j \neq k$, applying to individual r at calendar time t.

A.1.1 Premium Rates

See equation (6.19) computing the rate of premium at calendar time t in the multiple-state model in Section 6.5.

(a) As long as insurer knows the sub-population to which and individual belongs, equation (6.19) is valid.

(b) However, in our model (Chapter 7), we assume, that the insurer does not know to which sub-population each individual belongs. Therefore, we charge weighted average premium rates over all sub-populations as follows:

  (i) Assume that insurers' information is that all individuals are in state $i0$ at birth but that i is unknown.

(ii) Thus, we need further condition to ensure that premiums are unaffected by movements between states of health before an insured event occurs.

(iii) The premium rate at calendar time t will not depend on being in state ij, $a_r^{ij}(t) = a_r(t)$.

(iv) After birth, where we denote a set pairs of ij by C_r and the probability $P\left[I_r^{i0}(0) = 1\right]$ by p_{i0}, we can define martingales in our model as follows:

$$E\left[E\left[\int_0^t \sum_{ij\in C_r}\sum_{k:j\neq k} e^{-\delta s} A_r^{ijk}(s) dM_r^{ijk}(s)\middle| I_r^{i0}(0) = 1\right]\right] = 0. \tag{A.3}$$

(v) Since equation (A.3) is decomposed into equations (A.4) and (A.5) as follows: (A.3)=(A.4)−(A.5) in which:

$$E\left[E\left[\sum_{ij\in C_r}\sum_{k:j\neq k} A_r^{ijk}(s) dN_r^{ijk}(s)\middle| I_r^{i0}(0) = 1\right]\right] = \sum_{ij\in C_r}\sum_{k:j\neq k} p_{i0} A_r^{ijk}(s)\, {}_sp_{r,0}^{i0j}\mu_{r,s}^{ijk} ds, \tag{A.4}$$

and:

$$E\left[E\left[\sum_{ij\in C_r}\sum_{k:j\neq k} A_r^{ijk}(s) I_r^{ij}(s)\mu_s^{jk} ds\middle| I_r^{i0}(0) = 1\right]\right] = \sum_{ij\in C_r}\sum_{k:j\neq k} p_{i0} A_r^{ijk}(s)\, {}_sp_{r,0}^{i0j}\mu_{r,s}^{ijk} ds, \tag{A.5}$$

we can derive the premium rate function in any state in our model, independent on being state ij, at calendar time t as weighted average overall all sub-populations at calendar time t, associated with mortality rates and sum assured as follows:

$$a_r(t) = \frac{\sum_{ij\in C_r}\sum_{k:j\neq k} p_{i0} A_r^{ijk}(t)\, {}_tp_{r,0}^{i0j}\mu_{r,t}^{ijk}}{\sum_{ij\in C_r}\sum_{k:j\neq k} p_{i0t}\, p_{r,0}^{i0j}}, \tag{A.6}$$

and where $A_r^{ijk}(t) = 1$, see equation (7.1). Bear in mind that C_r is defined as an underwriting class at which insurance is purchased in Section 7.4.3. However, the actuarial mathematics represented in this appendix shows

C_r is also valid for any set of states whatsoever.

A.1.2 Policy Values

Section 6.6 describes policy values in two-state (Section 6.4) and multiple-state models (Section 6.5). For the multiple-state multiple-population model, denote V_t^{ij} to be the policy value at calendar time t in state ij (given that being in state ij at calendar time t):

$$V_t^{ij} = E[L_r^{ij}(t)|\mathcal{F}_t] = E\left[e^{\delta t}\int_t^\infty \sum_l e^{-\delta w} dL_r^{il}(w)\middle| I_r^{ij}(t) = 1\right]. \tag{A.7}$$

(See Section 6.6, $\mathcal{F}_t$ refers to the complete life history up to and including calendar time t.) These policy values can be calculated by solving Thiele's differential equation if the life history of a single individual is Markov. See also Thiele's differential equations and their numerical solutions in multiple-state models which are easily applied to the multiple-state multiple-population above.

A.1.3 A Measure of Adverse Selection Losses

We measure of adverse selection costs (Section 6.7) in such model (Section A.1):

(a) Suppose that the insurer set premiums $\tilde{a}_r(t)$, based on the information available about individual r.

(b) Then the insurer charges a premium rate $\tilde{a}_r(t)$ when $a_r(t)$ is the 'correct' premium rate associated with the true nature of individual r in which assume that $\tilde{a}_r(t) < a_r(t)$. In other words, we assume the insurers have less information than individuals under adverse selection.

(c) As a result, our measure of the adverse selection costs in such model is:

$$\frac{E\left[\int_0^\infty \sum_{ij\in C_r}\sum_{k:j\neq k} e^{-\delta t}\left(A_r^{ijk}(t)N_r^{ijk}(t) - \tilde{a}_r(t)I_r^{ij}(t)dt\right)\middle| I_r^{i0}(0) = 1\right]}{E\left[\int_0^\infty \sum_{ij\in C_r}\sum_{k:j\neq k} e^{-\delta t}\tilde{a}_r(t)I_r^{ij}(t)dt\middle| I_r^{i0}(0) = 1\right]}. \tag{A.8}$$

Bibliography

Adabag, A. S., Kuskowski, M. A. & Maron, B. J. (2006), 'Determinants for clinical diagnosis of hypertrophic cardiomyopathy', *The American Journal of Cardiology* **98**(11), 1507–1511.

AHA (2011), 'The normal heart and how it works', *American Heart Association* .

Arias, E., Heron, M. & Xu, J. (2017), 'United States Life Tables, 2013.', **66**(3).

Bick, A. G., Flannick, J., Ito, K., Cheng, S., Vasan, R. S., Parfenov, M. G., Herman, D. S., DePalma, S. R., Gupta, N., Gabriel, S. B., Funke, B. H., Rehm, H. L., Benjamin, E. J., Aragam, J., Taylor, H. A., Fox, E. R., Newton-Cheh, C., Kathiresan, S., O'Donnell, C. J., Wilson, J. G., Altshuler, D. M., Hirschhorn, J. N., Seidman, J. G. & Seidman, C. (2012), 'Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts', *The American Journal of Human Genetics* **91**(3), 513–519.

Brook, L., Lennard, F., Shenton, A., Lalloo, F., Ambus, I., Ardern-Jone, A., Belk, R., Kerr, B., Craufurd, D., Eeles, R. & Gareth Evans, D. (2004), 'BRCA1/2 predictive testing: A study of uptake in two centres.', *European Journal of Human Genetics* **12**, 654–662.

Cairns, A., Dickson, D., Macdonald, A., Waters, H. & Willder, M. (1998), 'Stochastic processes: Learning the language', *Faculty of Actuaries students' society* .

Charron, P., Carrier, L., Dubourg, O., Tesson, F., Desnos, M., Richard, P., Bonne, G., Guicheney, P., Hainque, B., Bouhour, J., Mallet, A., Feingold, J., Schwartz, K. & Komajda, M. (1997), 'Penetrance of familial hypertrophic cardiomyopathy.', *Genetic Counseling (Geneva, Switzerland)* **8**(2), 107–114.

Charron, P., Héron, D., Gargiulo, M., Richard, P., Dubourg, O., Desnos, M., Bouhour, J.-B., Feingold, J., Carrier, L., Hainque, B., Schwartz, K. & Komajda, M. (2002), 'Genetic testing and genetic counselling in hypertrophic cardiomyopathy: The French experience', *Journal of Medical Genetics* **39**(10), 741–746.

Christiaans, I., Birnie, E., Bonsel, G. J., Mannens, M. M., Michels, M., Majoor-Krakauer, D., Dooijes, D., van Tintelen, J. P., van den Berg, M. P., Volders, P. G., Arens, Y. H., van den Wijngaard, A., Atsma, D. E., Helderman-van den Enden, A. T., Houweling, A. C., de Boer, K., van der Smagt, J. J., Hauer, R. N., Marcelis, C. L., Timmermans, J., van Langen, I. M. & Wilde, A. A. (2011), 'Manifest disease, risk factors for sudden cardiac death, and cardiac events in a large nationwide cohort of predictively tested hypertrophic cardiomyopathy mutation carriers: Determining the best cardiological screening strategy', *European Heart Journal* **32**(9), 1161.

Christiaans, I., Birnie, E., Bonsel, G. J., Wilde, A. A. M. & van Langen, I. M. (2008), 'Uptake of genetic counselling and predictive DNA testing in hypertrophic cardiomyopathy', *European Journal of Human Genetics* **16**(9), 1201–07.

Christian, S., Atallah, J., Clegg, R., Giuffre, M., Huculak, C., Dzwiniel, T., Parboosingh, J., Taylor, S. & Somerville, M. (2018), 'Uptake of predictive genetic testing and cardiac evaluation for children at risk for an inherited arrhythmia or cardiomyopathy', *Journal of Genetic Counseling* **27**(1), 124–130.

Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2013), *Actuarial Mathematics for Life Contingent Risks*, International Series on Actuarial Science, second edn, Cambridge University Press.

Elliott, P., Andersson, B., Arbustini, E., Bilinska, Z., Cecchi, F., Charron, P., Dubourg, O., Kühl, U., Maisch, B., McKenna, W. J., Monserrat, L., Pankuweit, S., Rapezzi, C., Seferovic, P., Tavazzi, L. & Keren, A. (2008), 'Classification of the cardiomyopathies: A position statement from the European Society of Cardiology working group on myocardial and pericardial diseases', *European Heart Journal* **29**(2), 270–276.

Elliott, P. M., Anastasakis, A., Borger, M. A., Borggrefe, M., Cecchi, F., Charron, P., Hagege, A. A., Lafont, A., Limongelli, G., Mahrholdt, H., McKenna, W. J., Mogensen, J., Nihoyannopoulos, P., Nistri, S., Pieper, P. G., Pieske, B., Rapezzi, C., Rutten, F. H., Tillmanns, C. & Watkins, H. (2014), '2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy', *European Heart Journal* .

Elliott, P. M., Gimeno, J. R., Thaman, R., Shah, J., Ward, D., Dickie, S., Tome Esteban, M. T. & McKenna, W. J. (2006), 'Historical trends in reported survival rates in patients with hypertrophic cardiomyopathy', *Heart* **92**(6), 785–791.

Falconer, D. S. & Mackay, T. F. (1990), *Introduction to Quantitative Genetics*, fourth edn, Pearson Education Limited.

Geelen, E., Horstman, K., Marcelis, C. L., Doevendans, P. A. & Van Hoyweghen, I. (2012), 'Unravelling fears of genetic discrimination: An exploratory study of Dutch HCM families in an era of genetic non-discrimination acts.', *European Journal of Human Genetics* **20**, 1018–23.

Gersh, B. J., Maron, B. J., Bonow, R. O., Dearani, J. A., Fifer, M. A., Link, M. S., Naidu, S. S., Nishimura, R. A., Ommen, S. R., Rakowski, H., Seidman, C. E., Towbin, J. A., Udelson, J. E. & Yancy, C. W. (2011), '2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy', *Circulation* **124**(24), e783–e831.

Gui, E., Lu, B., Macdonald, A., Waters, H. & Wekwete, C. (2006), 'The genetics of breast and ovarian cancer iii: A new model of family history with insurance applications', *Scandinavian Actuarial Journal* **2006**(6), 338–367.

Gutiérrez, C. & Macdonald, A. S. (2004), 'Huntington's disease, critical illness insurance and life insurance', *Scandinavian Actuarial Journal* **2004**, 279–313.

Holloway, S., Bernhard, B., Campbell, H. & Lam, W. W. (2008), 'Uptake of testing for BRCA1/2 mutations in South East Scotland.', *European Journal of Human Genetics* **16**, 906–912.

Howard, R. C. W. (2014), 'Report to CIA research committee: Genetic testing model: If the underwriters had no access to known results', *Canadian Institute of Actuaries (CIA)* .

Husser, D., Ueberham, L., Jacob, J., Heuer, D., Riedel-Heller, S., Walker, J., Hindricks, G. & Bollmann, A. (2018), 'Prevalence of clinically apparent hypertrophic cardiomyopathy in Germany—An analysis of over 5 million patients', *PLOS One* **13**(2).

Jensen, M. K., Havndrup, O., Christiansen, M., Andersen, P. S., Diness, B., Axelsson, A., Skovby, F., Køber, L. & Bundgaard, H. (2013), 'Penetrance of hypertrophic cardiomyopathy in children and adolescents: A 12 year follow-up study of clinical screening and predictive genetic testing', *Circulation* **127**, 48–54.

Khouzam, A., Kwan, A., Baxter, S. & Bernstein, J. A. (2015), 'Factors associated with uptake of genetics services for hypertrophic cardiomyopathy', *Journal of Genetic Counseling* **24**(5), 797–809.

Khurana, I. (2014), *Textbook of Human Physiology for Dental Students*, Elsevier Health Sciences.

Lane, M., Ngueng Feze, I. & Joly, Y. (2015), 'Genetics and personal insurance: The perspectives of Canadian cancer genetic counselors', *Journal of Genetic Counseling* **24**, 1022–1036.

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

Lu, L., Macdonald, A. S., Waters, H. R. & Yu, F. (2007), 'A dynamic family history model of hereditary nonpolyposis colorectal cancer and critical illness insurance', *Annals of Actuarial Science* **2**(2), 289–325.

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

Macdonald, A. S., Richards, S. J. & Currie, I. D. (2018), *Modelling Mortality with Actuarial Applications*, International Series on Actuarial Science, Cambridge University Press.

Macdonald, A. S., Waters, H. R. & Wekwete, C. T. (2003*a*), 'The genetics of breast and ovarian cancer. I: A model of family history', *Scandinavian Actuarial Journal* **2003**, 1–27.

Macdonald, A. S., Waters, H. R. & Wekwete, C. T. (2003*b*), 'The genetics of breast and ovarian cancer. II: A model of critical illness insurance', *Scandinavian Actuarial Journal* **2003**, 28–50.

Macdonald, A. & Yu, F. (2011), 'The impact of genetic information on the insurance industry: Conclusions from the 'bottom-up'modelling programme', *Astin Bulletin* **41**(02), 343–376.

Maron, B. J. (2004), 'Hypertrophic cardiomyopathy: An important global disease', *The American Journal of Medicine* **116**(1), 63–65.

Maron, B. J., Casey, S. A., Poliac, L. C., Gohman, T. E., Almquist, A. K. & Aeppli, D. M. (1999), 'Clinical course of hypertrophic cardiomyopathy in a regional United States cohort', *JAMA* **281**(7), 650–655.

Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T. & Bild, D. E. (1995), 'Prevalence of hypertrophic cardiomyopathy in a general population of young adults echocardiographic analysis of 4111 subjects in the CARDIA study', *Circulation* **92**(4), 785–789.

Maron, B. J., Olivotto, I., Spirito, P., Casey, S. A., Bellone, P., Gohman, T. E., Graham, K. J., Burton, D. A. & Cecchi, F. (2000), 'Epidemiology of hypertrophic cardiomyopathy–related death', *Circulation* **102**(8), 858–864.

Maron, B. J., Ommen, S. R., Semsarian, C., Spirito, P., Olivotto, I. & Maron, M. S. (2014), 'Hypertrophic cardiomyopathy: Present and future, with translation into contemporary cardiovascular medicine', *Journal of the American College of Cardiology* **64**(1), 83–99.

Maron, B. J., Roberts, W. C. & Epstein, S. E. (1982), 'Sudden death in hypertrophic cardiomyopathy: A profile of 78 patients.', *Circulation* **65**(7), 1388–1394.

Maron, B. J., Rowin, E. J., Casey, S. A., Haas, T. S., Chan, R. H., Udelson, J. E., Garberich, R. F., Lesser, J. R., Appelbaum, E., Manning, W. J. & Maron, M. S. (2013), 'Risk stratification and outcome of patients with hypertrophic cardiomyopathy ≥60 years of age', *Circulation* **127**(5), 585–593.

Maron, B. J., Rowin, E. J., Casey, S. A., Lesser, J. R., Garberich, R. F., McGriff, D. M. & Maron, M. S. (2016a), 'Hypertrophic cardiomyopathy in children, adolescents and young adults associated with low cardiovascular mortality with contemporary management strategies', *Circulation* **133**, 62–73.

Maron, B. J., Rowin, E. J., Casey, S. A., Link, M. S., Lesser, J. R., Chan, R. H., Garberich, R. F., Udelson, J. E. & Maron, M. S. (2015), 'Hypertrophic cardiomyopathy in adulthood associated with low cardiovascular mortality with contemporary management strategies', *Journal of the American College of Cardiology* **65**(18), 1915–1928.

Maron, B. J., Towbin, J. A., Thiene, G., Antzelevitch, C., Corrado, D., Arnett, D., Moss, A. J., Seidman, C. E. & Young, J. B. (2006), 'Contemporary definitions and classification of the cardiomyopathies: An American heart association scientific statement from the council on clinical cardiology, heart failure and transplantation committee; quality of care and outcomes research and functional genomics and translational biology interdisciplinary working groups; and council on epidemiology and prevention', *Circulation* **113**(14), 1807–1816.

Maron, M. S., Hellawell, J. L., Lucove, J. C., Farzaneh-Far, R. & Olivotto, I. (2016b), 'Occurrence of clinically diagnosed hypertrophic cardiomyopathy in the United States', *The American Journal of Cardiology* **117**(10), 1651–1654.

Martin, J. A., Hamilton, B. E., Osterman, M. J. & Driscoll, A. K. (2018), 'Births: Final data for 2017.', **67**(8).

McKenna, W. J. & Elliott, P. (2015), 'Live with hypertrophic cardiomyopathy', *British Heart Foundation* .

Meijers-Heijboer, E., Verhoog, L., Brekelmans, C., Seynaeve, C., Tilanus-Linthorst, M., Wagner, A., Dukel, L., Devilee, P., van den Ouweland, A., van Geel, A. & Klijn, J. (2000), 'Presymptomatic DNA testing and prophylactic surgery in families with a BRCA1 or BRCA2 mutation', *The Lancet* **355**(9220), 2015–2020.

Miller, E. M., Wang, Y. & Ware, S. M. (2013), 'Uptake of cardiac screening and genetic testing among hypertrophic and dilated cardiomyopathy families', *Journal of Genetic Counseling* **22**(2), 258–267.

Niimura, H., Bachinski, L. L., Sangwatanaroj, S., Watkins, H., Chudley, A. E., McKenna, W., Kristinsson, A., Roberts, R., Sole, M., Maron, B. J., Seidman, J., Seidman, C. E., Thierfelder, L., Jarcho, J. A., Anastasakis, A., Toutouzas, P., Elstein, E., Liew, C.-C., Liew, J., Mably, J., Rakowski, H., Wigle, E. D., Zhao, M., Salerni, R. & Bjornsdottir, H. (1998), 'Mutations in the gene for cardiac myosin-binding protein c and late-onset familial hypertrophic cardiomyopathy', *New England Journal of Medicine* **338**(18), 1248–1257.

Norberg, R. (1991), 'Reserves in life and pension insurance', *Scandinavian Actuarial Journal* **1991**(1), 3–24.

Norberg, R. (1992), 'Hattendorff's theorem and Thiele's differential equation generalized', *Scandinavian Actuarial Journal* **1992**(1), 2–14.

Office for National Statistics (2018), 'Births in England and Wales: 2017', Office for National Statistics, UK.

Page, S. P., Kounas, S., Syrris, P., Christiansen, M., Frank-Hansen, R., Andersen, P. S., Elliott, P. M. & McKenna, W. J. (2012), 'Cardiac myosin binding protein-c mutations in families with hypertrophic cardiomyopathy: Disease expression in relation to age, gender, and long term outcome', *Circulation: Cardiovascular Genetics* **5**(2), 156–166.

Pokorski, R. J. (2002), 'Hypertrophic cardiomyopathy: Risk factors for life and living benefits insurance.', *Journal of Insurance Medicine* **34**, 43–60.

Prince, A. E. R. (2019), 'Political economy, stakeholder voices, and saliency: Lessons from international policies regulating insurer use of genetic information', *Journal of Law and the Biosciences* **5**(3), 461–494.

Ropka, M. E., Wenzel, J., Phillips, E. K., Siadaty, M. & Philbrick, J. T. (2006), 'Uptake rates for breast cancer genetic testing: A systematic review', *Cancer Epidemiology and Prevention Biomarkers* **15**(5), 840–855.

Sanz, J., Ramón y Cajal, T., Torres, A., Darder, E., Gadea, N., Velasco, A., Fortuny, D., López, C., Fisas, D., Brunet, J., Alonso, M. C. & Balmaña, J. (2010), 'Uptake of predictive testing among relatives of BRCA1 and BRCA2 families: A multicenter study in Northeastern Spain', *Familial Cancer* **9**(3), 297–304.

Seidman, J. & Seidman, C. (2001), 'The genetic basis for cardiomyopathy: From mutation identification to mechanistic paradigms', *Cell* **104**(4), 557–567.

Semsarian, C., Ingles, J., Maron, M. S. & Maron, B. J. (2015), 'New perspectives on the prevalence of hypertrophic cardiomyopathy', *Journal of the American College of Cardiology* **65**(12), 1249–1254.

Sharaf, R. N., Myer, P., Stave, C. D., Diamond, L. C. & Ladabaum, U. (2013), 'Uptake of genetic testing by relatives of Lynch syndrome probands: A systematic review', *Clinical Gastroenterology and Hepatology* **11**, 1093–1100.

Spirito, P., Bellone, P., Harris, K. M., Bernabò, P., Bruzzi, P. & Maron, B. J. (2000), 'Magnitude of left ventricular hypertrophy and risk of sudden death in hypertrophic cardiomyopathy', *New England Journal of Medicine* **342**(24), 1778–1785.

Subramanian, K., Lemaire, J., Hershey, J. C., Pauly, M. V., Armstrong, K. & Asch, D. A. (1999), 'Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance', *The Journal of Risk and Insurance* **66**(4), 531–550.

Sudbery, P. (2002), *Human Molecular Genetics*, second edn, Pearson Education Limited.

Suthers, G. K., Armstrong, J., McCormack, J. & Trott, D. (2006), 'Letting the family know: Balancing ethics and effectiveness when notifying relatives about genetic testing for a familial disorder', *Journal of Medical Genetics* **43**(8), 665–670.

Terauchi, Y., Kubo, T., Baba, Y., Hirota, T., Tanioka, K., Yamasaki, N., Furuno, T. & Kitaoka, H. (2015), 'Gender differences in the clinical features of hypertrophic cardiomyopathy caused by cardiac myosin-binding protein c gene mutations', *Journal of Cardiology* **65**(5), 423–428.

Thomas, R. G. (2012), 'Genetics and insurance in the United Kingdom 1995–2010: The rise and fall of "scientific" discrimination', *New Genetics and Society* **2012**, 1–20.

Van Driest, S. L., Vasile, V. C., Ommen, S. R., Will, M. L., Tajik, A. J., Gersh, B. J. & Ackerman, M. J. (2004), 'Myosin binding protein c mutations and compound heterozygosity in hypertrophic cardiomyopathy', *Journal of the American College of Cardiology* **44**(9), 1903–10.

Wagner, A., Tops, C., Wijnen, J. T., Zwinderman, K., van der Meer, C., Kets, M., Niermeijer, M. F., Klijn, J. G. M., Tibben, A., Vasen, H. F. A. & Meijers-Heijboer, H. (2002), 'Genetic testing in hereditary non-polyposis colorectal cancer families with a MSH2, MLH1, or MSH6 mutation', *Journal of Medical Genetics* **39**(11), 833–837.

Watkins, H., McKenna, W. J., Thierfelder, L., Suk, H. J., Anan, R., O'Donoghue, A., Spirito, P., Matsumori, A., Moravec, C. S., Seidman, J. & Seidman, C. E. (1995), 'Mutations in the genes for cardiac troponin t and α-tropomyosin in hypertrophic cardiomyopathy', *New England Journal of Medicine* **332**(16), 1058–1065.

Whitaker, R. H. (2006), 'Anatomy of the heart', *Medicine* **34**(5), 163–165.

Wordsworth, S., Leal, J., Blair, E., Legood, R., Thomson, K., Seller, A., Taylor, J. & Watkins, H. (2010), 'DNA testing for hypertrophic cardiomyopathy: A cost-effectiveness model', *European Heart Journal* **31**(8), 926–935.