GENETICS, INSURANCE AND HYPERTROPHIC CARDIOMYOPATHY

By Oytun Haçariz†, Torsten Kleinow†and Angus S. Macdonald†

ABSTRACT

We specify a mathematical model of Hypertrophic Cardiomyopathy (HCM) and consider the potential costs arising from adverse selection in a life insurance market. HCM is a dominantly inherited heart disorder which is relatively common and has high mortality; a population prevalence of 0.2% and annual mortality hazard (force of mortality) of 1% have been widely cited. Adverse selection may arise if insurers may not take account of adverse DNA-based genetic tests for a causative mutation. A novel feature of our model is that it includes cascade genetic testing (CGT) in nuclear families. CGT is the form of testing used in HCM. Among other things, it implies that genetic testing occurs only if a family history exists. We find in most scenarios the premium increases necessitated by adverse selection to be very small — fractions of 1% — but we ask under what circumstances these might increase to appreciable levels. Insurers' inability to use family history in underwriting would have a large impact. We note that the epidemiology of HCM is still evolving and that 0.2% is likely to be a considerable underestimate of mutation prevalence, while recent estimates of annual mortality hazards are much less than 1%. The first of these in particular is likely to limit any adverse selection costs.

KEYWORDS

Adverse Selection, Genetic Epidemiology, Cascade Genetic Testing, Hypertrophic Cardiomyopathy, Life Insurance, Mortality

CONTACT ADDRESS

† Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK, and the Maxwell Institute for Mathematical Sciences.
Corresponding author: Angus Macdonald, A.S.Macdonald@hw.ac.uk.

1. Introduction

It is more than twenty years since the advent of genetic testing based on sequencing DNA, bringing with it concerns about how this information may be used, that are even now not resolved. Life and health insurance has been one of the most contentious issues.

At the heart of this debate has been the perception that genetic test results have extraordinary predictive power, compared with other medical information that an insurance underwriter may use. In a few cases this is true, which is perhaps unfortunate, as most of the time the truth is far more nuanced.

Macdonald & Yu (2011) summarized a programme of research into the key question, namely, what costs might insurers face from adverse selection if they were unable to access genetic test results or, more generally, family medical history? They found the costs to be small, of the order of 1% of premiums, chiefly due to the rarity of the conditions considered. A later study by Howard (2014) found costs in Canada to be much higher, of the order of 10% of premiums or more, chiefly because he included a number of relatively common

heart conditions not considered by Macdonald & Yu (2011). Subsequently Lombardo (2018) repeated Howard's analysis for the USA.

Our aim in this paper is to model in detail one of these heart conditions, hypertrophic cardiomyopathy (HCM), and its impact on a life insurance market. The general methodology we apply has been described elsewhere (Subramanian et al. (1999), Macdonald et al. (2003*a*), Macdonald et al. (2003*b*), Gui et al. (2006), Lu et al. (2007)) but to make this paper more self-contained we outline it in Appendix A. We add one new feature in this paper, namely modelling cascade genetic testing within families (see Section 3).

We find that the progress of HCM, from the onset of clinical effects and possible symptoms, to the endpoint of death from various causes, is very nuanced indeed. While genetic testing is now part of clinical practice, the interpretation of test results, in terms of individual risk, is far from clear-cut; in fact as time passes it seems that their predictive power may be less than once thought. The impact of HCM on an insurance market, consequently, covers a wide range of possible outcomes, of which the most extreme may be thought to be rather unlikely.

2. Hypertrophic Cardiomyopathy (HCM)

This section summarizes a more detailed account in the PhD thesis of Haçarız (Haçarız 2020). Basic references are Elliott et al. (2014) and Gersh et al. (2011).

2.1 *Clinical Onset and Prognosis*

Hypertrophic cardiomyopathy (HCM) is the thickening, over time, of the muscle wall of the left ventricle of the heart. Once the wall thickness reaches a certain threshold (13 to 15 millimetres in adults, depending on the standard used) the clinical onset of HCM is deemed to have occurred. In a high proportion of cases this happens before ages 20–30. The main causes of death below age 60 are sudden cardiac arrest (SCA) and heart failure (HF). Above age 60 stroke is also a significant cause of death. HCM appears to affect males and females equally.

It is important to distinguish between *onset* of HCM, which happens when the left ventricular wall thickness reaches a certain threshold, and *clinical diagnosis* of HCM, which means its presence is detected. There is strong evidence that most clinical HCM is 'silent', meaning present but never detected (see Section 2.2.1).

Left ventricular wall thickness is measured physically by means of an echocardiogram.

An affected person may be completely asymptomatic, or may display various stages of heart failure. HCM is one of several causes of death by SCA of fit young people, including sometimes well-known athletes.

2.2 *The Genetics of HCM*

HCM is caused by mutations in any one of several genes encoding sarcomere proteins[1]. All are dominantly inherited[2], so inheriting a mutation from just one parent is sufficient

[1] Long fibrous proteins with a rôle in the beating of the heart muscle.

[2] With certain exceptions, every individual carries two copies of each gene, in general different variants, each inherited with probability 1/2 from the two variants carried by each of their parents. In turn the individual will pass on one or other of their two gene variants to each child they have, independently, with probability 1/2. This expresses Mendel's laws.

to confer the full risk of HCM. Unlike some disorders, such as breast cancer, of which a modest proportion is inherited, HCM is always associated with a heritable mutation. From Elliott et al. (2014), ignoring unrelated mutations, mutations accounting for about 2/3 to 3/4 of clinical HCM have been identified.

Two figures are widely cited in the secondary literature: (a) HCM affects about 0.2% of the population; and (b) the mortality hazard among affected persons is about 1% per year. The third basic parameter is the *penetrance*, which is the probability that a mutation carrier actually develops the disorder[3]. We deal with each of these in turn.

2.2.1 *Prevalence*

The prevalence of 0.2% is based on Maron et al. (1995), in which seven out of 4,111 unselected subjects were found to have clinical HCM. Several other studies (see Semsarian et al. (2015)) reached similar conclusions. However, many fewer patients are diagnosed with HCM than this prevalence would imply. This observation led to two very large recent studies. Maron et al. (2016b), found 59,009 HCM-related claims in a medical database with 169,098,614 patients (more than half of the US population), a prevalence of about 0.035%. Husser et al. (2018) found 4,000 cases among 5,490,810 patients in a German healthcare claims database, a prevalence of about 0.07%.

There is an even more important way in which HCM may be 'silent'. Bick et al. (2012) sequenced the DNA of 3,600 unselected subjects for known HCM-related mutations and found them in 22 cases, a prevalence of 0.6%. Only four of these 22 subjects had clinical HCM. Bearing in mind that not all HCM-related mutations have been identified (see Section 2.2 above), the prevalence of mutations might be as high as 1%.

In other words, the prevalence of relevant mutations is much higher than the prevalence of clinically-present HCM, which in turn is much higher than the prevalence of clinically-detected HCM. The first of these matters if we suppose that persons might choose to buy life insurance on the basis of a genetic test result.

2.2.2 *Mortality*

The mortality hazard of 1% per year also has to be interpreted with care. It, and previous estimates, appear to be based on clinical studies in which the endpoints used to calculate the 'mortality' hazard are not necessarily fatal. For example, all studies we have seen use 'Sudden Cardiac Death' as an endpoint, whether or not the subject died. 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. (It is for this reason that we prefer to use the term 'Sudden Cardiac Arrest'.) For actuarial use, it is clearly necessary to remove these non-fatal 'deaths' from the mortality hazard. Doing so based on three recent large studies

[3]Variations in DNA define the genotype. The corresponding variations in physical expression define the range of *phenotypes*.

(Maron et al. (2013), Maron et al. (2015) and Maron et al. (2016a)) we estimate a hazard of 0.55% per year at all relevant ages. Note that this hazard rate is conditional on the presence of clinical HCM, not the presence of an HCM-related mutation. The details are in Haçarız (2020).

2.2.3 *Penetrance*

The penetrance of a mutation is the probability that the associated phenotype (in this case the thickened left ventricular wall) is present. In the case of a condition which develops after birth, penetrance is a function of age, which we denote by $F(x)$, and the lifetime penetrance is the probability that the associated phenotype should ever develop during life. The lifetime penetrance of HCM has been estimated to be 69% (Charron et al. 1997) but this must be qualified.

(a) HCM-related mutations have variable expression. Mutations in different genes, and different mutations in the same gene, may have different penetrances, see Section 4.2(b).
(b) If there is a high proportion of 'silent' mutations (see Section 2.2.1) penetrance estimates based on clinically-detected HCM may be overstated. The consequences for adverse selection are significant and are discussed in Section 7.1.

3. Cascade Genetic Testing

Population-based screening is rarely advocated for genetic disorders. One reason is cost; most genetic conditions are rare. Another is the difficulty of assessing risk on the basis of DNA sequence alone. Ancillary information, such as a history of the disorder in blood relatives, is often needed to assess risk and advise patients, see Dunn et al. (2013).

Cascade genetic testing (CGT) overcomes these problems and is the form of testing used in HCM (Elliott et al. 2014). It proceeds as follows.

(a) Clinical HCM is identified in a member of a family in which HCM has not previously manifested. This person is called the *proband* or the index case.
(b) The proband is tested for the presence of a known HCM-related mutation (see Section 2.2).
(c) If a known mutation is found, genetic testing is offered to first-degree relatives[4] of the proband. They are not obliged to take it up. Any who test negative are presumed to be not at risk. Any who test positive are screened for clinical HCM, possibly for life, and offered suitable treatment.
(d) If no known mutation is found in the proband, all first-degree relatives of the proband are offered screening for clinical HCM, possibly for life, and offered suitable treatment.
(e) Depending on the results for individual first-degree relatives of the proband, testing/screening may be offered to second-degree relatives of the proband, and so on. In this way, the testing can 'cascade' through several branches of a family.

CGT has several implications for adverse selection in life insurance. First, no-one is tested, or even offered testing, without the existence of an affected family member.

[4]First-degree relatives are parents, siblings and children.

Thus, a family history is always present. Second, given a family with n members, at most $n-1$ could ever buy insurance knowing they are at risk of HCM (since someone must become a proband first) and Mendel's laws reduce the maximum to $(n-1)/2$ on average (see footnote 2). If n is not large, the pool of potential 'adverse selectors' is reduced considerably. Third, the extent of CGT depends on the willingness of relatives of the proband to take up the offer.

These features shape the information available to individuals, to guide their insurance purchasing decisions, and to insurers, to set premium rates. Therefore it is important to model CGT.

4. The Model

4.1 *Events and Information*

For our purposes, a model of a life insurance market must reflect three key features.

(a) It must represent the whole population able to participate in the market (by choosing to buy insurance or not).
(b) It must represent the biological risks within that population (sub-populations with different genotypes having different risks of disease and death).
(c) It must represent information available to individuals and insurers that inform their decisions, respectively, whether or not take a genetic test and/or buy insurance, and what premiums to charge for it. In particular, it must represent the information revealed by HCM-related events and CGT.

We broadly follow the methodology used before in Subramanian et al. (1999), Macdonald et al. (2003*a*), Macdonald et al. (2003*b*), Gui et al. (2006) and Lu et al. (2007) (see Appendix A), which was to specify a Markov model in continuous time, in which both biological events (movements between states of health, conditioned on genotypes) and individuals' behaviour (decision to have a genetic test, decision to buy life insurance) are represented by transitions between states. The state space is partitioned into sub-populations representing all possible genotypes so the entire population is modelled, with occupancy probabilities in the 'starting' states in each sub-population given by the genotype prevalences. The information available to individuals and to insurers at any time is represented by knowing that the individual is present in one of a set of states, but not necessarily knowing exactly which state.

Figure 1 shows a model of the life history of an individual with a given HCM genotype, including biological events and decisions made by the individual. Non-fatal HCM events are included because of the information they may reveal: that there is a mutation in the family. There are separate non-fatal HCM states for persons who are 'tested-insured' and 'untested-insured' because they may be paying different rates of premium (see Section 4.6). There are no transitions from non-fatal HCM states into insured states because we assume such a transition would be medically underwritten and would not lead to adverse selection.

Each intensity into a fatal or non-fatal HCM state is a product of penetrance $F(x)$ and the intensity of the event conditional on clinical HCM having developed (see Section

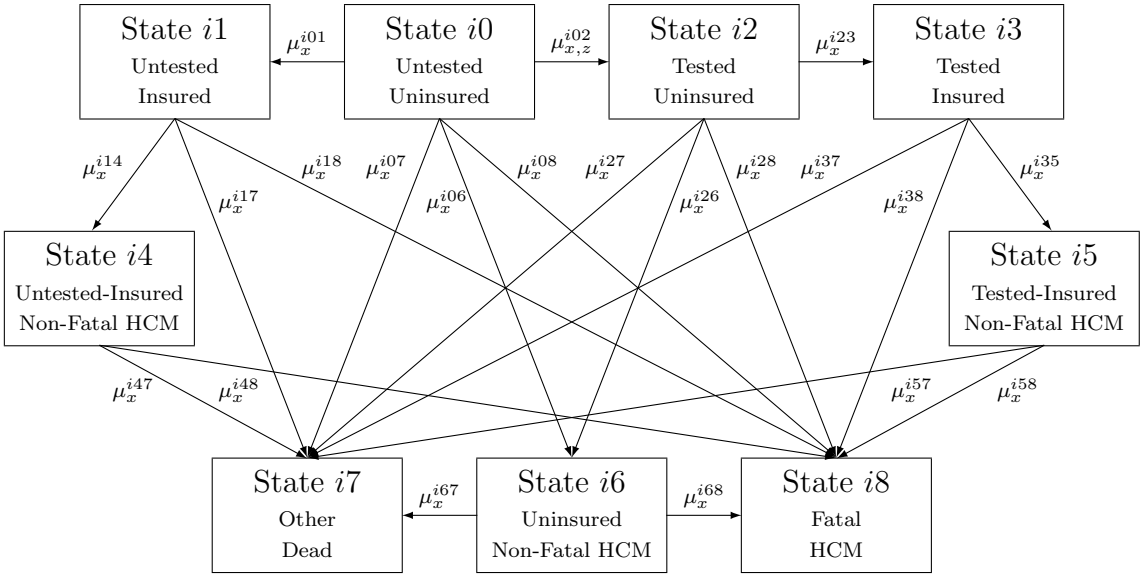

Figure 1: A mathematical model of the life history of a person in the ith of several subpopulations defined by HCM genotype. The life history includes biological events and decisions made by the individual. In $\mu_{x,z}^{i02}$, z refers to duration in state $i0$ since (if) a proband appeared in the family.

2.2.2). For example, if ρ_x^{i08} is the intensity of fatal HCM, conditional on clinical HCM being present, then $\mu_x^{i08} = F(x)\,\rho_x^{i08}$ and so on.

This framework is adequate as long as *all* material events can be represented as transitions within the model in Figure 1. This is not the case if we must model CGT, as information may be gained from events befalling other family members. In particular, the intensity of moving from an untested state to a tested state increases through CGT if a family member is diagnosed with HCM. (We could in theory expand the state space without limit, but this becomes very unwieldy.) We therefore adopt a two-pronged approach.

(a) The great majority of persons are in families with no HCM-related mutations present. We call these 'non-HCM families'. For them, HCM can never happen[5], CGT can never happen and the problem mentioned above does not exist. No information external to the model in Figure 1 arises, and we can calculate occupancy probabilities and expected present values (EPVs) by solving the Kolmogorov equations and Thiele's equations respectively.

(b) We call families in which an HCM-related mutation is present 'HCM families'. We deal with these by simulating their joint life histories (that is, the life histories of all the family members simultaneously in calendar time). CGT can be modelled explicitly within each family. So for these families, occupancy probabilities and expected present

[5]Strictly, an *ab initio* mutation could arise in the germline of a parent, which could then be passed on to their children, but this is rare and we ignore it.

values are calculated by Monte-Carlo estimation.

Sometimes we are forced to choose model parameters in the absence of reliable studies. In such cases, we generally choose conservatively, meaning that the choice would tend to increase any adverse selection costs suffered by insurers.

4.2 *Biological Parameters*

The biological parameters are: population mortality rates; population genotype prevalences; rates of onset of HCM; and rates of HCM-related events (SCA, heart failure, stroke) subdivided into fatal and non-fatal events.

(a) We assume that population mortality is that of the Life Tables, United States (US), 2013 (Arias et al. 2017), males and females.

(b) As a baseline we assume that the prevalence of HCM-related mutations is 0.2% (see Section 2.2.1). This is conservative. Later, we will consider the implications of the Bick et al. (2012) study which suggests prevalence exceeding 0.6%. We distinguish two types of HCM-related mutations, based on penetrance. Early-onset mutations are those where onset of clinical HCM often occurs in early adulthood; late-onset mutations are those where onset is typically later than this. Based on Elliott et al. (2014) we assume that 75% of mutations are early-onset and 25% are late-onset. We also assume that 2/3 of each type are known and 1/3 are unknown (see Section 2.2). We therefore assume there are four HCM sub-populations defined by the genotypes early-onset/known, early-onset/unknown, late-onset/known and late-onset/unknown.

(c) (1) There are no reliable studies of the age-related penetrance of early-onset mutations, partly because of the difficulty of research involving children. We make the very conservative assumption that everyone with an early-onset mutation who survives to age 20 has suffered onset of clinical HCM; that is, $F(20) = 1$.

(2) There are two reliable studies of penetrance in respect of mutations in the most common gene we have classed as late-onset: one Dutch (Christiaans et al. 2011) and one Japanese (Terauchi et al. 2015). We use the first of these and note that sensitivity testing in Haçarız (2020) shows that the choice makes almost no difference. The estimated age-related penetrances, for males and females, are stated in Appendix B.

(d) We noted in Section 2.2.2 our estimated HCM-related mortality hazard rate of 0.0055 per year. Hazard rates of non-fatal HCM events are strongly age-related. We need them because they can signal the presence of a mutation in a previously unaffected family, thus they can initiate CGT. Based on the studies referred to in Section 2.2.2, we assume annual hazard rates of 0.009509 at ages 0–29, 0.003472 at ages 30–59, and 0.001611 at ages 60 and over.

(e) Since we are interested only in ages up to 60, we neglect stroke as an HCM-related cause of death (see Haçarız (2020)).

For members of non-HCM families, who are not exposed to the risk of HCM, only the population mortality matters.

4.3 *Simulation of Life Histories and Family Formation*

The population consists of independent families. A family consists of two parents and a number of children (possibly zero). We assume that all parents are born at time $t = 0$ and all children are born at time $t = 30$. Because the *family* model includes individuals of different ages, it is based on calendar time $t \geq 0$. However, the transition intensities in Figure 1 are functions of age (with one exception, see Section 4.4(b)). Thus, at calendar time $t \geq 0$ parents are age t, and at calendar time $t \geq 30$ children are age $t - 30$, and the intensities are functions of these ages. In detail, family histories are simulated as follows.

(a) A large number of persons are born at time $t = 0$ with HCM-related mutations, distributed among four sub-populations according to the assumed prevalences (see Section 4.2(b)). A proportionately large number of persons are born with no HCM-related mutations.
(b) The life history of each person is simulated up to time $t = 20$, using the model in Figure 1 and a sufficiently small time-step.
(c) At time $t = 20$ each surviving person in the HCM population is given a spouse of the opposite sex from the non-HCM population. We ignore the possibility that both spouses have an HCM-related mutation.
(d) The *joint* life history of each person in the HCM population *and their spouse* is simulated up to time $t = 30$, using the model in Figure 1 for each individual, and a sufficiently small time-step.
(e) At time $t = 30$ each spouse-pair in which both are alive has a number S of children, where S is a Poisson random variable with parameter λ (thus $S = 0$ is possible). For both UK and US populations, $\lambda \approx 1.8$, which we use as a baseline (Office for National Statistics (2018), Martin et al. (2018)).
(f) Children are assigned to either sex with probability 1/2, and to a genotype according to Mendel's law (so they each, independently, inherit the mutation with probability 1/2, see footnote 2.
(g) After time $t = 30$ an HCM family consists of a parent from the HCM population (the carrier parent), a spouse from the non-HCM population (the non-carrier parent) and S children of which some might be mutation carriers.
(h) The *joint* life history of each person in the HCM population and their spouse (if alive) and children (if born) is simulated up to time $t = 90$, using the model in Figure 1 for each individual, and a sufficiently small time-step. Any life insurance the parents have expires at time $t = 60$.
(i) At time $t = 90$ all children are age 60 and, in our model, any life insurance they have expires.

Thus, all life histories are modelled in continuous time, with specific events external to the life history model in Figure 1 taking place at times $t = 0, 20$ and 30. Each life history can be regarded as a marked point process with certain marks being acquired at predetermined times.

4.4 *Cascade Genetic Testing*

If a proband appears in a family, CGT is initiated. Therefore the simulation of life histories described in Section 4.3 may include relevant events (transitions) as follows.

(a) If an HCM-related event (fatal or non-fatal) occurs in the family for the first time,

the affected person is the proband. This can happen before or after children are born, and the proband can be the carrier parent or any carrier child. The proband is assumed to undergo genetic testing and, if this identifies a known mutation, CGT is initiated.

(b) If CGT is initiated, we assume that testing is offered to all at-risk family members. The outcome is represented by a constant transition intensity of 0.6931472 between 'untested' states and the corresponding 'tested' states, operating for one year only. Thus, after one year, each person offered testing has accepted the offer with probability approximately 1/2. This is based on the available literature on the uptake of testing in HCM and other inherited diseases, summarized in Haçarız (2020). We consider more extensive testing in Section 6.5. We assume that family members who do not accept testing in the first year will not be tested.

(c) If the testing in (a) above does not identify a known mutation, each at-risk family member is assumed to carry the mutation with probability 1/2. This is justified for children of a carrier parent by Mendel's law, and for the parents of a carrier child by the assumption that only one of the parents carries a mutation.

The above is a brief description from which some second-order details have been omitted (for example, what happens if a parent becomes a proband while the children are very young, or before they are born?). Full details and references can be found in Haçarız (2020).

The outcome is that, at any calendar time t, we know the exact status of each family member (age, sex, genotype, tested/untested, affected/unaffected by HCM, alive/dead) but the individuals themselves may have more limited knowledge (in respect of genotype) and an insurer more limited knowledge still.

In the next sections, we describe how we model the combination of biological risks, information and individual decisions (see also Appendix A).

4.5 *Sub-populations*

Figure 2 shows the full model with nine sub-populations, each one a copy of the model in Figure 1 with appropriate transition intensities.

(a) Since we have five distinct genotypes (non-carriers and carriers of early/late-onset and known/unknown mutations) the presence of five of these sub-populations in the model is self-explanatory; those labelled $i = 0, 2, 4, 6, 8$. The idea of a person being chosen at random from the general population is that they are allocated at birth to one of the starting states $00, 20, 40, 60$ or 80 with probability equal to the respective genotype prevalences (see Section 4.2(b)).

(b) The rôle of the other four sub-populations, those labelled $i = 1, 3, 5, 7$, is less obvious. They are needed to represent individuals who are: (i) members of a family in which one of the four HCM genotypes is present (meaning that one parent is a carrier), but; (ii) are not carriers themselves. Individuals are *not* allocated at random to these four sub-populations in the same way as in (a) above. These sub-populations contain only spouses and children of mutation carriers, as follows.

  (1) When spouse-pairs are formed at time $t = 20$ (see Section 4.3(c)) the spouse of each mutation carrier is moved from sub-population 0 to the sub-population

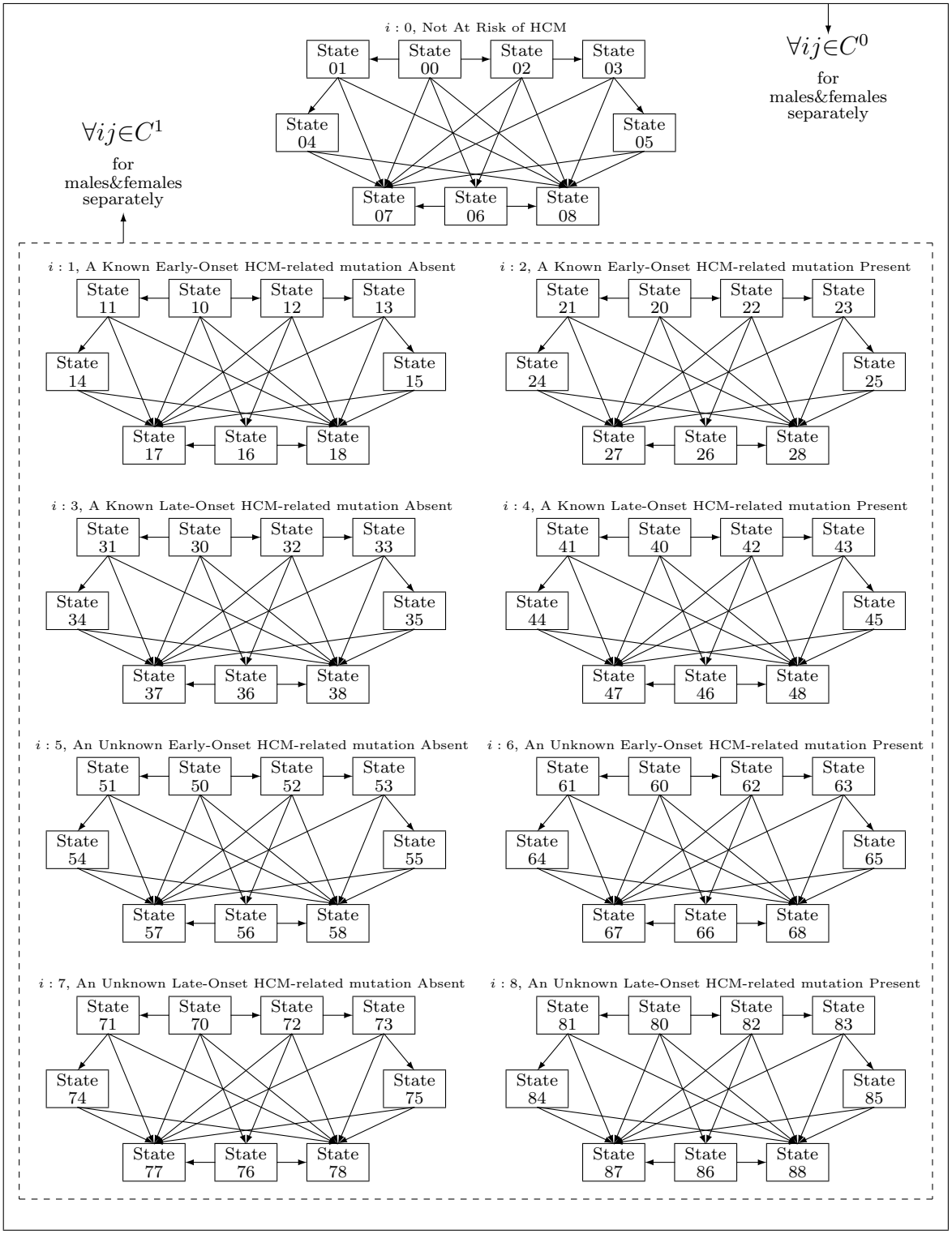

Figure 2: A mathematical model of the life history of a randomly chosen individual, present in one of the states $i0$ at birth with probabilities determined by HCM-related mutation prevalences and Mendel's laws. See Section 4.5, and Appendix A.

$i = 1, 3, 5$ or 7 corresponding to their partner's sub-population.

(2) When children are born at time $t = 30$ they are allocated at random, with probability 1/2, either to the same sub-population as their carrier parent, or to the corresponding non-carrier sub-population (see Section 4.3(f)).

The biological risks of individuals in sub-populations 1, 3, 5 and 7 are the same as those of individuals in sub-population 0; they are all non-carriers. Other transition intensities may be different, however, because of the information known to the individual and to the insurer, in particular, family history and genetic test results. This is why we need these four sub-populations in the model.

4.6 *Insurers' Premium Rating Decisions*

As in the works cited in Section 1, individuals may purchase life insurance at any age between 20 and 60, and if they do the policy is in force until age 60. Insurers do not charge level premiums depending on the age at purchase, but age-dependent premium rates equal (in principle) to the hazard rate of death. This is similar to the system of mortality charges often used for unit-linked business in the UK.

Suppose, for simplicity, that insurers never know genetic test results. They may know family medical history, which in our model means knowing whether or not there is a proband in an applicant's family. This divides the population into two underwriting classes.

(a) If there is a proband in an applicant's family, the insurer knows they are a member of an HCM family, but not which sub-population (early/late onset, known/unknown mutation) nor whether the applicant actually is a carrier or not. This collection of sub-populations is indicated by the dashed box in Figure 2, where this underwriting class is labelled C^1. However the insurer is capable of calculating the probabilities of all these cases, as occupancy probabilities in our model.

(b) If there is no proband in an applicant's family, the insurer knows nothing. The applicant could be a member of an HCM family, or a member of a non-HCM family, in each case with probabilities that the insurer again can calculate. This collection of sub-populations is indicated by the solid box in Figure 2, where this underwriting class is labelled C^0.

The insurer acquires no new information after insurance is purchased, except in the event of death. Persons continue to be charged premium rates for the underwriting class at the time of purchase. At every time t after purchase, the insurer knows: (a) a set of states in which the applicant could be; (b) the probability that the applicant is in each of these states; and (c) hazard rates of death out of each of these states. The insurer will never know which state the person is actually in. It is intuitive (and is shown in Haçarız (2020)) that the appropriate age-related rate of premium per unit of benefit is the occupancy-probability-weighted average of these hazard rates of death. (A very small allowance must be made for the prevalence of mutations falling in the childrens' generation, because of HCM-related deaths before age 30 in the parents' generation. The details are in Haçarız (2020).)

Contrary to the above, our default assumption will be that insurers may be informed of negative genetic test results, that show conclusively that an individual does not carry

a mutation. This is: (a) consistent with the moratorium in force in the UK; and (b) generally conservative, as these individuals can then be offered normal premium rates. We achieve this simply (though approximately) by charging individuals in the relevant insured states inside the dashed box in Figure 2 the premium rates for underwriting class C^0 rather than those for underwriting class C^1.

We also consider, in Sections 5 et seq., the possibility that insurers are banned from using family history as well as genetic tests, in which case there is only the one, C^0, underwriting class.

4.7 *Individuals' Insurance Purchase Decisions*

We model individuals' insurance purchase decisions as transition intensities between 'uninsured' states and the corresponding 'insured' states. The transition intensities depend on the information available to the individual. For this purpose the individual is allocated, at any time, to one of four 'information classes', analogous to the insurer's two underwriting classes. We define these as follows.

(i) Class ζ^n. There is no proband in the individual's family. Members of non-HCM families are always in this information class. Members of HCM families are in this information class as long as no proband exists in the family.

(ii) Class ζ^{50}. There is a proband in the individual's family. The individual has not been tested, either by choice or because the proband carries an unknown mutation, so knows only that they are at 50% risk of carrying the mutation.

(iii) Class ζ^{100}. There is a proband in the individual's family, carrying a known mutation, the individual has been tested and knows that they carry the mutation.

(iv) Class ζ^0. There is a proband in the individual's family, carrying a known mutation, the individual has been tested and knows that they do not carry the mutation.

If the proband is the carrier parent, their spouse remains in information class ζ^n, but if the proband is a carrier child, both parents are in information class ζ^{50} until (if) being tested, or the carrier parent suffering an HCM-related event.

In the various scenarios that we model, the insurance purchasing decisions are modelled by specifying the rates of insurance purchase as functions of the information class. This can extend to the amount of insurance purchased too.

4.8 *An Example of Different Levels of Information*

Suppose a child is in sub-population $i = 1$; a member of a family in which one parent carries a known early-onset mutation, but not themselves a carrier. Suppose they reach age 20 (become economically active) with no proband in the family (no carrier parent or child has yet suffered an HCM event). Suppose also that insurers are allowed to use family history but not genetic test results.

(a) As long as no proband has appeared, neither the individual nor the insurer knows which sub-population they are in. They could be in any uninsured and untested healthy state, $00, 10, 20, \ldots, 80$. The individual is in information class ζ^n and underwriting class C^0. Under these circumstances no adverse selection is possible.

(b) When they are age 25, a proband appears in their family (a parent or a sibling). The proband is tested and found to be in sub-population $i = 2$. As long as the individual is not tested, they know only that they are in one of states 10 or 20. The insurer knows only that the individual is in one of states $10, 20, \ldots, 80$ *or* one of states $12, 22, \ldots, 82$ (the insurer does not know if the individual has been tested, and now that a proband has appeared that is a possibility). The individual is in information class ζ^{50} and underwriting class C^1.
(c) The individual chooses to be tested and finds out that they are not a mutation carrier. They must be in state 10. But the information available to the insurer has not changed. The individual is in information class ζ^0 and underwriting class C^1.

If insurers were allowed to use negative genetic test results (which will in fact be our baseline assumption), then at stage (c) above the insurer would know the same as the individual, and the latter would be in information class ζ^0 and underwriting class C^0.

4.9 *Overview of the Modelling Approach*

Consider a portfolio of life insurance contracts. Define B to be the discounted benefit outgo, and P to be the discounted premium income. Then $L = B - P$ is the loss under this block of business. Under the Principle of Equivalence, the insurer sets premium rates so that $\mathrm{E}[L] = 0$.

Suppose now that the population is heterogeneous, consisting of group X, with loss L_X such that $\mathrm{E}[L_X] > 0$, and group Y with loss L_Y such that $\mathrm{E}[L_Y] = -\mathrm{E}[L_X] < 0$. Suppose that group X gains an informational advantage over the insurer, such that some of its members know or believe that $\mathrm{E}[L_X] > 0$, but the insurer does not know this. If the incentives are strong enough, group X can increase its purchases of insurance which increases its discounted benefits to B_X^*, its discounted premiums proportionately to $P_X^* = P_X B_X^*/B_X$ and its expected loss to $\mathrm{E}[B_X^* - P_X^*] = \mathrm{E}[L_X^*] > \mathrm{E}[L_X]$. The total expected loss is now $\mathrm{E}[L_X^* + L_Y] > 0$. To eliminate this expected loss, while being unable to pass the cost on to group X alone, the insurer would have to increase all premiums by a proportion:

$$\frac{\mathrm{E}[L_X^* + L_Y]}{\mathrm{E}[P_X^* + P_Y]}. \tag{1}$$

This is our measure of the cost of adverse selection. Group X consists of HCM families. Group Y consists of non-HCM families. Calculating the measure in equation (1) requires us to compute:

(a) $\mathrm{E}[L_X^*]$ and $\mathrm{E}[P_X^*]$, which we compute by Monte-Carlo simulation of family life histories (see Section 4.4); and
(b) $\mathrm{E}[L_Y]$ and $\mathrm{E}[P_Y]$, which we compute by solving Thiele's equations numerically.

Not everyone in group X either is or believes themselves to be at increased risk, only those in information classes ζ^{50} and ζ^{100}. Nevertheless, we must model entire families as explained in Section 4.4.

Information class ζ^{50} deserves special attention. If insurers are allowed to use family history, then any increase in insurance purchasing by this group will increase $\mathrm{E}[P_X^*]$ but not $\mathrm{E}[L_X^*]$. Essentially, the insurer is able to charge increased costs to those who give rise

to them *if* it is allowed to identify them. The residual increased costs, which the insurer cannot allocate to identifed individuals, are spread over everyone. This will be clear in the subsequent tables, in which we assume that insurers may or may not use family history.

5. Baseline Assumptions for Adverse Selection Costs

5.1 *Process*

Establishing a baseline for measuring the costs of adverse selection follows two stages.

(a) We first define a model of a life insurance market in which no adverse selection exists. This corresponds to the case $\mathrm{E}[L_X + L_Y] = 0$ in Section 4.9.

(b) We then vary certain parameters to introduce a degree of adverse selection. This corresponds to the case $\mathrm{E}[L_X^* + L_Y] > 0$ in Section 4.9.

In fact our baseline assumptions under adverse selection define not one scenario but four, since we always consider two levels of adverse selection ('mild' and 'severe') and the two cases that insurers may or may not use family history.

We divide the model parameters into two groups, biological and behavioural.

5.2 *Biological Parameters*

What we called 'biological parameters' were set out in Section 4.2. Two of these will be of particular interest, in the light of Sections 2.2.1 and 2.2.2.

(a) We noted in Section 2.2.1 the difference between the population prevalence of clinical HCM of about 0.2%, and the population prevalence of *known* HCM-related mutations of about 0.6%. We take the latter to imply a population prevalence of all HCM-related mutations of about 0.9% (extrapolating from 0.6% in Bick et al. (2012) allowing for 1/3 of mutations being unknown).

We choose the first of these as our baseline, because the figure of 0.2% is so widely cited. However, we shall return to this question in Section 7.1, because 0.9% is more consistent with the trajectory of genetic epidemiology, and has considerable implications for adverse selection costs.

(b) We noted in Section 2.2.2 that the endpoints used in the epidemiological literature on HCM often include both fatal and non-fatal HCM events. The widely-cited annual mortality hazard rate of 0.01 may be a case in point. We choose as our baseline the more reliable annual mortality hazard rate of 0.0055, and age-dependent hazards of non-fatal HCM events, given in Section 4.2.

In Section 4.3(e) the number of children in a family was assumed to have a Poisson distribution with parameter $\lambda = 1.8$. This is a fairly robust baseline, based on US and UK populations. Larger values of λ may be used as a proxy, at least approximately, for CGT spreading outside the modelled nuclear family, and testing being offered to a larger number of individuals, following the discovery of a single proband. See Section 6.5.

Biological parameters are the same in the model with no adverse selection, and in the baseline model with adverse selection, the two stages mentioned in Section 5.1 above.

5.3 *Behavioural Parameters*

What we call the 'behavioural parameters' are those concerning the uptake of genetic testing under CGT, the premium rates charged by insurers, and the insurance purchasing behaviour of individuals. All of these depend on the information available: (a) uptake of testing on the existence of a proband (Section 4.4); (b) premium rating on the allowable underwriting classes (Section 4.6); and (c) insurance purchasing behaviour on the information classes (Section 4.7).

5.3.1 *No Adverse Selection*

The key assumption in calculating expected present values of benefit and premium cashflows in the absence of adverse selection is the 'normal' rate at which individuals purchase insurance between ages 20 and 60. That is, the transition intensity between uninsured and insured states. Our baseline is an annual intensity of 0.05, in every subpopulation, representing a large insurance market. Later, an annual intensity of 0.01 will represent a much smaller insurance market. (These were the rates used in the previous works cited in Section 1. They correspond roughly to 5% or 1%, respectively, of uninsured individuals buying insurance in any year.) We assume that all sums assured are £1, and that there is no genetic testing at all.

Note that we do not attempt to model a competitive market with insurers competing on price. By 'market' we mean the aggregate demand for insurance, modelled by the intensities of purchasing insurance, and we simply assume that insurers collectively supply the quantity of insurance that is demanded.

We consider two cases in all circumstances; either insurers are allowed to use family history, or they are not. In either case, they charge the correct premiums calculated as in Section 4.6, using the correct occupancy probabilities, so the measure in equation (1) is zero in the absence of adverse selection.

5.3.2 *Baseline Cases of Adverse Selection*

(a) From Section 4.4(b), we assume that any individual offered genetic testing upon the appearance of a proband in the family accepts it within one year with probability 1/2. Some second-order assumptions, such as children being born when a proband exists, are described fully in Haçarız (2020) and are omitted here.
(b) From Section 4.6 insurers charge different premium rates to individuals who have a family history of HCM (a proband in the family) and those who do not, with one exception: individuals who have tested negative for a known mutation, who must have a family history of HCM, are treated as if they did not.
(c) Baseline levels of adverse selection are defined by individuals in information classes ζ^{100} and ζ^{50} (Section 4.7) purchasing insurance at annual rate 0.10 (twice normal, called 'mild' adverse selection) or at annual rate 0.25 (five times normal, called 'severe' adverse selection). These assumptions are based on Macdonald & Yu (2011) and the sources cited therein.

5.4 *Baseline Adverse Selection: Results*

Table 1 shows the percentage premium increases due to adverse selection in the four baseline scenarios. Also shown are 95% quantile intervals. These refer to the distributions

Table 1: Percentage increases in premiums caused by adverse selection under baseline adverse selection scenarios (Section 5.3.2). The 95% quantile intervals are in respect of the Monte-Carlo estimation of mean EPVs of cashflows in the HCM families.

Adverse	Mean Premium Increase and 95% QI			
Selection	Family History Disallowed		Family History Allowed	
	%	%	%	%
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)

generated by Monte-Carlo estimation of $\mathrm{E}[L_X^*]$ and $\mathrm{E}[P_X^*]$ as in Section 4.9. We observe the following.

(a) The premium increases are very small, fractions of one percent.
(b) Inability to use family history makes a great difference, increasing the premium increases by a factor of about 2.7.
(c) The premium increases with no adverse selection are zero to four decimal places, although some elements of equation (1) are calculated by Monte-Carlo simulation (this will be the case in subsequent tables as well). This is an artefact arising because the small adjustment to premium rates mentioned in Section 4.6 must be made after the simulations are carried out. Hence we refer to 'quantile intervals' and not 'confidence intervals'.

We can partition the loss L_X^* in equation (1) into the contributions from each of the four types of HCM family (early/late onset, known/unknown mutations). With family history not allowed, late-onset mutations account for only 0.978% and 0.864% of the premium increases, under 'mild' and 'severe' adverse selection respectively. With family history allowed, late-onset mutations actually reduce premium increases, and the corresponding contributions are −6.908% and −7.298%. The prevalences of early- and late-onset mutations are clearly significant for the costs of adverse selection, a point that we discuss further in Section 7.2.

We now consider factors that would amplify the adverse selection losses (Section 6) or that would diminish the adverse selection losses (Section 7) and then consider what conclusions to draw from the whole range of possibilities (Section 8).

6. What Factors May Amplify Adverse Selection?

6.1 *Comparison with Howard (2014)*

Our baseline adverse selection costs are very small. Most of the behavioural assumptions are conjectural, and the biological assumptions, while based on epidemiology, may lie within some plausible range. Varying the assumptions may increase or decrease the costs. In this section we consider factors that may increase the costs.

One reason for doing so is the disparity between the results of Macdonald & Yu (2011) and Howard (2014). By far the largest part of the adverse selection costs in the latter were

Table 2: Percentage increases in premiums caused by adverse selection in a smaller life insurance market (Section 6.3).

Purchase Intensity	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
		%	%	%	%
0.05	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
0.05	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
0.05	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
0.01	None	0.0000	(−0.0411 , 0.0411)	0.0000	(−0.0416 , 0.0408)
0.01	Mild	0.0431	(−0.0005 , 0.0837)	0.0149	(−0.0275 , 0.0568)
0.01	Severe	0.3225	(0.2636 , 0.3810)	0.1143	(0.0571 , 0.1708)

accounted for by inherited heart disorders, which were not considered by Macdonald & Yu (2011). Direct comparison of the two models is not possible, but it would be useful to estimate, approximately, what percentage premium increases were attributable to HCM alone in Howard (2014). Then we could ask by how much the parameters in our model would have to change if the costs attributable to HCM were to approach those in Howard (2014).

In Howard (2014) the 'cost sub-model' showed that total premium income in Canada, estimated at $3.5 billion in 2012, would have to rise by $405,455,952 to cover the cost of adverse selection, a rise of about 12%. HCM accounted for $89,187,658 of the increase, or about 22%. (It was the second highest single contributor to the losses. For comparison, inherited breast cancer contributed just $5,363,834, about 1.3% of the increase.) That suggests that an increase in total premiums of about 2.5% from HCM in our model is roughly equivalent.

6.2 *Unavailability of Family History*

We have already noted that inability to use family history in underwriting amplifies adverse selection losses considerably, by a factor of about 2.7 in the baseline scenarios. Nevertheless, the percentage losses are still very small.

6.3 *A Smaller Life Insurance Market*

Our baseline rate of insurance purchase is a constant 0.05 per year, chosen to represent a large, mature life insurance market. To represent a smaller market we use a rate of insurance purchase of 0.01 per year. 'Mild' adverse selection is represented by double the normal purchase rate (0.02) and 'severe' adverse selection by an extreme purchase rate again of 0.25. The results are shown in Table 2.

The costs are increased by a factor of about 1.6 with mild adverse selection and about 6 with severe adverse selection.

Our simple assumption of a lower transition intensity into insured states results in a much smaller proportion of individuals in non-HCM families having insurance in force at any time. A similar effect would be achieved by having a higher rate of insurance purchase combined with a substantial lapse rate. Howard (2014) assumed a lapse rate of

3% per annum, falling to 0.5% per annum for 'substandard' lives, essentially those who knew they were at risk of HCM. For that reason it is possible that Table 2 gives a fairer comparison with Howard (2014) than Table 1.

6.4 *Higher Sums Insured*

Macdonald & Yu (2011) noted the observation in Gutiérrez & Macdonald (2004), that if adverse selection meant individuals took out larger sums insured, the resulting premium increases were almost proportionate to those with normal sums insured. The reason was that model states were partitioned in advance into three underwriting classes, within each of which the expected loss was zero under no adverse selection. That is, under no adverse selection there were expected losses of $\mathrm{E}[L_X^{(1)}] = -\mathrm{E}[L_Y^{(1)}]$, $\mathrm{E}[L_X^{(2)}] = -\mathrm{E}[L_Y^{(2)}]$ and $\mathrm{E}[L_X^{(3)}] = -\mathrm{E}[L_Y^{(3)}]$, adapting the notation of equation (1). Adverse selection affected either one or two of these underwriting classes. The effect of larger sums insured could therefore be gauged, approximately, as a multiple of the results with unit sum insured in Macdonald & Yu (2011).

Here matters are different. It is not possible to partition states into fixed underwriting classes in advance because the underwriting classes depend on whether there is a proband, and if so, who it is. The most that we can say is that the loss L_X^* and premium income P_X^* in equation (1) may be decomposed into $L_X^{*(1)}$ and $P_X^{*(1)}$ in respect of individuals in those information classes in which more insurance is purchased, and $L_X^{*(2)}$ and $P_X^{*(2)}$ in respect of others. If individuals in the first group then purchase a multiple n times the normal sum insured, premiums increase by a proportion:

$$\frac{\mathrm{E}[nL_X^{*(1)} + L_X^{*(2)} + L_Y]}{\mathrm{E}[nP_X^{*(1)} + P_X^{*(2)} + P_Y]}. \tag{2}$$

For large n the numerator is dominated by its first term, but for smaller n (say $n \leq 2$) it need not be, because L_Y can be quite large and is always negative. The denominator is dominated by its last term.

Howard (2014) assumed that the normal sum insured was \$100,000, and that adverse selectors increased this to \$1,000,000, that is, by a factor of ten. In Table 3 we show explicitly the effect of adverse selectors taking out sums insured of 2, 4 and 10 times the average. We see that the premium increases increase more than in proportion to the sum insured, as we should expect from equation (2).

6.5 *More Genetic Testing*

Adverse selection opportunities would arise more often if more individuals took genetic tests. We consider two ways in which this could happen.

(a) More people offered genetic testing could accept it. We assumed that half of those offered testing accepted it within one year of it being offered, and there was no testing thereafter (see Section 4.4(b)). As an alternative, we increase the annual intensity of transferring from an untested to a tested state to 4.60517, so that approximately 99% of individuals are tested within one year. The results are shown in Table 4. There is practically no difference when family history is disallowed, because we have assumed that persons in information classes ζ^{100} and ζ^{50} behave identically. Otherwise, the difference is not very large.

Table 3: Percentage increases in premiums caused by increased sums insured under adverse selection (Section 6.4). The column headed 'Adverse Selection' refers to the rate of insurance purchase.

Sum Insured	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
		%	%	%	%
1 ×	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
1 ×	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
1 ×	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
2 ×	None	0.0555	(0.0254 , 0.0863)	0.0188	(−0.0117 , 0.0488)
2 ×	Mild	0.1073	(0.0782 , 0.1405)	0.0375	(0.0076 , 0.0714)
2 ×	Severe	0.1655	(0.1323 , 0.2001)	0.0584	(0.0246 , 0.0922)
4 ×	None	0.1665	(0.1220 , 0.2123)	0.0564	(0.0121 , 0.1023)
4 ×	Mild	0.2699	(0.2210 , 0.3213)	0.0937	(0.0483 , 0.1436)
4 ×	Severe	0.3859	(0.3325 , 0.4485)	0.1352	(0.0837 , 0.1957)
10 ×	None	0.4981	(0.4062 , 0.6020)	0.1683	(0.0788 , 0.2628)
10 ×	Mild	0.7549	(0.6431 , 0.8653)	0.2604	(0.1518 , 0.3686)
10 ×	Severe	1.0429	(0.9080 , 1.1917)	0.3628	(0.2259 , 0.5113)

Table 4: Percentage increases in premiums due to a higher take-up rate of genetic testing, with approximately 99% acceptance instead of 50% (Section 6.5(a)).

Percentage Tested	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
%		%	%	%	%
50	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
50	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
50	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
99	None	0.0000	(−0.0240 , 0.0245)	0.0000	(−0.0247 , 0.0237)
99	Mild	0.0259	(0.0022 , 0.0498)	0.0131	(−0.0110 , 0.0377)
99	Severe	0.0551	(0.0297 , 0.0779)	0.0276	(0.0017 , 0.0507)

Table 5: Percentage increases in premiums due to higher values of λ, the mean number of children. This is a proxy for CGT spreading outside the nuclear family (Section 6.5(b)).

λ	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
		%	%	%	%
1.8	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
1.8	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
1.8	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
3.0	None	0.0000	(−0.0196 , 0.0201)	0.0000	(−0.0198 , 0.0195)
3.0	Mild	0.0372	(0.0153 , 0.0591)	0.0133	(−0.0088 , 0.0356)
3.0	Severe	0.0796	(0.0573 , 0.1033)	0.0288	(0.0069 , 0.0525)
5.0	None	0.0000	(−0.0179 , 0.0185)	0.0000	(−0.0179 , 0.0178)
5.0	Mild	0.0517	(0.0340 , 0.0704)	0.0186	(0.0005 , 0.0377)
5.0	Severe	0.1108	(0.0909 , 0.1320)	0.0404	(0.0212 , 0.0623)
7.0	None	0.0000	(−0.0147 , 0.0148)	0.0000	(−0.0153 , 0.0144)
7.0	Mild	0.0631	(0.0474 , 0.0799)	0.0228	(0.0075 , 0.0399)
7.0	Severe	0.1349	(0.1160 , 0.1536)	0.0494	(0.0294 , 0.0672)

(b) CGT could spread beyond the nuclear family. We cannot model this directly, but as a proxy we can increase the parameter λ, the mean number of children in a family (see Section 4.3(e)). This has the effect that a single proband exposes more individuals to the offer of genetic testing. Except for the fact that all the extra individuals are of the same (childrens') generation, the results should be plausible, because it is a feature of CGT that anyone who is offered genetic testing, even in the furthest reaches of an extended family, is a mutation carrier with probability 1/2.
The results are shown in Table 5, for $\lambda = 1.8, 3.0, 5.0$ and 7.0. The last of these approximates to testing 'cascading' from the proband's family into roughly three other related nuclear families, and increases the adverse selection costs by a factor of about 2.5.

However, we should note that more genetic testing would bring more individuals into treatment for HCM and, if effective, this would reduce mortality. See Section 7.5.

6.6 *Higher HCM-Related Mortality*

An HCM-related annual mortality hazard rate of 0.01 seems to be widely cited, although not supported by more recent research unless 'non-fatal deaths' are treated as actual deaths (see Section 2.2.2). Nevertheless, since it is so widespread, we show in Table 6 the effect of using an annual HCM-related mortality hazard rate of 0.01, and a hazard rate of non-fatal HCM-related events of zero. If family history is allowed in underwriting, the difference is negligible, for two reasons.

(a) We assume the insurer knows the 'correct' hazard, 0.01 or 0.0055 as the case may be, in calculating premiums in the C^1 underwriting class.

(b) The combined hazards of fatal and non-fatal HCM events (see Section 4.2(d)) are somewhat higher than 0.01 before age 29, and somewhat lower than 0.01 from age

Table 6: Percentage increases in premiums with HCM-related annual mortality hazard rates of 0.0055 and 0.01 (Section 6.6).

HCM-related Hazard Rate	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
		%	%	%	%
0.0055	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
0.0055	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
0.0055	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
0.01	None	0.0000	(−0.0240 , 0.0251)	0.0000	(−0.0236 , 0.0249)
0.01	Mild	0.0301	(0.0028 , 0.0565)	0.0095	(−0.0164 , 0.0353)
0.01	Severe	0.0643	(0.0338 , 0.0936)	0.0207	(−0.0087 , 0.0484)

30, and either type of event will initiate CGT.

If family history is not allowed in underwriting, the increase is still small, a consequence of the fact that even with a mortality rate as extreme as $q_x = 0.01$, more individuals survive than die over long periods (for example, $0.99^{40} \approx 2/3$, while $0.9945^{40} \approx 4/5$). The behavioural parameters are more influential than this biological parameter.

7. What Factors May Diminish Adverse Selection?

7.1 *Higher Mutation Prevalences*

We noted in Section 2.2.1 that the prevalence of known HCM-related mutations was about three times higher than the prevalence of clinical HCM. This has significant implications for the modelling assumption that insurance-purchasing behaviour is affected by genetic test results. The incidence of HCM-related events in the population may be taken to be a fixed datum, arising from clinical observations. So far we have assumed that such incidence is linked to a mutation prevalence of 0.2%, and the transition intensities of HCM-related events in Section 2.2.2 are consistent with that. If the actual mutation prevalence is 0.9% (see Section 5.2(a)) there are two extreme possibilities.

(a) The prevalence of 'active' HCM-related mutations is 0.2% and something causes the remainder to be 'silent' in families that carry them. We need not suppose that 'active' and 'silent' mutations are distinct; some other genetic or environmental factor could cause the same mutation to be 'active' in some families and 'silent' in others. In this case CGT will never be initiated in families carrying a 'silent' mutation and our previous modelling is unaffected. Other forms of genetic testing, however, such as screening or whole-genome sequencing at birth, would reveal the 'silent' mutations and the interpretation of HCM risk would not be possible on the basis of genotype alone.

(b) The prevalence of 'active' HCM-related mutations is 0.9% but their penetrances are much lower than estimated to date, if we are to reproduce the observed incidence of HCM-related events. The fact that penetrance might have been overestimated in the

Table 7: Percentage increases in premiums with prevalence of HCM-related mutations 0.2% and 0.9% (Section 7.1).

Mutation Prevalence	Adverse Selection	Mean Premium Increase and 95% QI			
		Family History Disallowed		Family History Allowed	
%		%	%	%	%
0.2	None	0.0000	(−0.0239 , 0.0245)	0.0000	(−0.0250 , 0.0242)
0.2	Mild	0.0259	(0.0025 , 0.0503)	0.0094	(−0.0145 , 0.0349)
0.2	Severe	0.0550	(0.0294 , 0.0786)	0.0198	(−0.0071 , 0.0436)
0.9	None	0.0000	(−0.0394 , 0.0391)	0.0000	(−0.0392 , 0.0388)
0.9	Mild	0.0090	(−0.0300 , 0.0479)	0.0029	(−0.0358 , 0.0419)
0.9	Severe	0.0189	(−0.0205 , 0.0582)	0.0062	(−0.0324 , 0.0453)

past would be the result of *selection bias*: it has been estimated in populations in which clinical HCM is present, not populations in which HCM-related mutations are present. See Section 7.4.

It is possible that insurers, like epidemiologists, see a selected population of HCM families, because HCM comes to their attention for a reason. However, unlike epidemiologists, insurers are not actively looking for HCM patients and families, so the reasons that they come to the attention of insurers and epidemiologists might be quite different.

To model the scenario in (b) above, we note that the transition intensities of interest to us (those in Figure 1) are a product of penetrance $F(x)$ and hazard rate conditional on clinical HCM being present, see Section 4.1. In the epidemiological literature, hazard rates of HCM-related events are conditioned on the presence of clinical HCM. Therefore we can attribute the higher prevalence to reduced penetrance. Our assumptions about penetrance were given in Section 4.2(c): we assumed $F(20) = 1$ for early-onset mutations and the penetrances of Christiaans et al. (2011) for late-onset mutations, see Appendix B. We calculate reduced hazard rates of HCM events by multiplying these penetrances by $0.22 \approx 0.2/0.9$. This is crude (but no more so than the conservative assumption that $F(20) = 1$ for early-onset mutations). The results are shown in Table 7. The adverse selection costs are reduced by a factor of about three.

7.2 *The Proportion of Late-Onset Mutations*

We noted in Section 2.2 that mutations accounting for about 2/3 to 3/4 of clinical HCM have been identified; in Section 4.2 that we assumed 25% of mutations to be late-onset; and in Section 5.4 that late-onset mutations accounted for a very small proportion of the adverse selection costs (less than 1% in the baseline case). We extrapolated the 25% proportion of late-onset mutations to those that are as yet unidentified. If that should be an underestimate, adverse selection costs would be lower than we have portrayed them.

7.3 *The Manner of Cascade Testing for Clinical HCM*

In the context of genetics and insurance, HCM (and other cardiomyopathies) differ from the classical single-gene disorders previously studied, except Adult Polycystic Kid-

ney Disease (APKD). Clinically-present HCM is a pre-existing condition in underwriting terms.

We assume that when a proband carrying a known mutation appears in a family, all relevant relatives are offered genetic testing and may then purchase insurance. An insurer will learn about the family history, if its use is permitted, but our model does not allow for the underwriter acquiring any other information, that could reveal HCM as a pre-existing condition. This could include diagnosed HCM, or symptoms without a diagnosis. 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.

The rôle of genetic counselors is important as long as genetic testing takes place in a clinical setting. Lane et al. (2015) said "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". If all patients received and followed such advice, the assumptions in our model would be plausible.

7.4 *Allowance for Selection Bias*

Studies such as this rely on the genetic epidemiological literature. This evolves. Typically, early studies identifying the genetic basis of a disease are based on small numbers of affected individuals or families and high mortality rates are observed. Over time, larger numbers of less affected, but still selected, individuals or families are studied and estimated mortality rates drop. Population-based studies lacking any selection bias are difficult to carry out, for rare disorders. Still rarer are studies based on underlying genotype rather than some aspect of the phenotype.

HCM is a case in point. Haçarız (2020) describes how studies from 1958 until the 1990s estimated annual mortality hazard rates of 3% to 6%. Since then opinion seems to have favoured the widely-cited annual mortality hazard rate of about 1%, but the most recent studies cited in Section 2.2.2 suggest mortality of about half that level. We have drawn attention in Section 7.1 to the much higher prevalence of mutations than clinically-present HCM (the only statistics cited in this paper that are based on unselected populations). Bick et al. (2012) noted that only four of 22 mutation carriers in their study had clinical HCM and said "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).

Epidemiologists have long been aware of *ascertainment bias*, namely that bias arising because families and individuals are studied only because they have come to the attention of the researcher. See Hodge (2002) and the references therein for the large literature on this subject. Lemaire et al. (2000) said, in their study of breast and ovarian cancer:

> "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

Table 8: Percentage increases in premiums when the insurer calculates premiums in the C^1 underwriting class assuming an annual hazard rate of fatal HCM of 0.01, and no non-fatal HCM events, when then actual mortality hazard is 0.0055, and the hazard rates of non-fatal HCM events are as in Section 4.2(d) (Section 7.4).

Sum Insured	Purchase Intensity	Adverse Selection	Mean Premium Increase and 95% QI			
			Family History Disallowed		Family History Allowed	
			%	%	%	%
1 ×	0.05	None	−0.2132	(−0.2369 , −0.1887)	−0.2357	(−0.2616 , −0.2121)
1 ×	0.05	Mild	−0.1873	(−0.2106 , −0.1629)	−0.2404	(−0.2646 , −0.2144)
1 ×	0.05	Severe	−0.1583	(−0.1838 , −0.1347)	−0.2460	(−0.2726 , −0.2210)
1 ×	0.01	None	−0.1911	(−0.2322 , −0.1502)	−0.2107	(−0.2517 , −0.1700)
1 ×	0.01	Mild	−0.1481	(−0.1916 , −0.1076)	−0.2197	(−0.2624 , −0.1770)
1 ×	0.01	Severe	0.1306	(0.0718 , 0.1889)	−0.2739	(−0.3330 , −0.2160)
10 ×	0.01	None	0.2697	(0.1003 , 0.4367)	−0.3114	(−0.4724 , −0.1511)
10 ×	0.01	Mild	0.6947	(0.4695 , 0.9037)	−0.3984	(−0.6080 , −0.1732)
10 ×	0.01	Severe	3.3937	(3.0155 , 3.8090)	−0.8991	(−1.3240 , −0.4820)

> 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".

Macdonald et al. (2003*b*), also in a study of breast and ovarian cancer, suggested reducing estimated onset transition intensities by as much as 50% or 75% to allow for this. We have not done so in this study.

As an example of the impact of selection bias, we suppose that the insurer calculates premiums in the C^1 underwriting class assuming an annual hazard rate of fatal HCM of 0.01, and no non-fatal HCM events, when the actual mortality hazard is 0.0055, and the hazard rates of non-fatal HCM events are as in Section 4.2(d). Table 8 shows the effect in the large and small markets when 'adverse selectors' purchase the normal sum insured (compare with Table 2), and in the small market when 'adverse selectors' purchase ten times the normal sum insured (the latter being associated with the highest adverse selection costs, see Section 8.2).

The results are striking, especially when family history is allowed. The selection bias introduces such a level of conservatism to the premium rates in underwriting class C^1 that the costs even of severe adverse selection may be completely overturned.

Bearing in mind all the sources of potential bias, and the typical trajectory of epidemiological studies, we are inclined to regard every value for premium increases in this paper as likely to be an overstatement.

7.5 *Improved Treatment of HCM*

Although not part of our model, it is probably true that treatment of HCM will continue to improve and that mortality will be further reduced. Even ignoring future developments, better detection of clinical HCM would bring treatment to more people

and hence reduce mortality. If that were the case, the premium increases shown in Table 5 might be overstated.

8. Discussion and Conclusions

8.1 *The Epidemiology of HCM*

HCM differs from most of the 'classical' genetic disorders studied in the actuarial literature (such as Huntington disease and inherited cancers) in that the phenotype develops early in life in a large proportion of cases, and is detectable by conventional means. If so detected, it would be underwritten as a pre-existing condition. In order for a person to exploit the informational advantage of a known genetic test result, they would have to know that test result in isolation, unaware of any other indication, disclosable to an insurer, that HCM might be present. As long as genetic testing for seriously pathogenic mutations happens in a clinical setting, this might limit the scope for large-scale adverse selection. Lane et al. (2015) makes it clear that testing is almost always in the context of a family history being present (Section 7.3).

If genetic testing moves to a wider setting, such as whole-genome sequencing at birth, the interpretation of a particular mutation being present changes radically, especially since it seems possible that the mutation prevalence is much higher than 0.2%, perhaps approaching 1%, and the penetrance is correspondingly reduced (see the quotation from Bick et al. (2012) in Section 7.4). This makes the insurance purchasing decision, based on a genetic test result, less well-determined.

(a) Individuals could continue to over-insure as envisaged in worst-case scenarios. Persons who will die because of HCM then join the insurance pool at less than cost, but a much larger number of persons who will not die because of HCM also join the insurance pool at more than cost, and the cost of adverse selection is significantly reduced.
(b) The much-reduced information content of a genetic test result, in the absence of ancillary information of a non-genetic nature, could diminish any incentive to over-insure. In particular it might diminish any incentive to stake a large quantity of money on insurance premiums in a gamble, the outcome of which looks less attractive.

The epidemiology of HCM is still evolving and, apart from some studies of prevalence, all of it is based on selected populations. Epidemiologists are well aware of the biases this introduces and actuarial studies should take this into account as well.

It is still too early to say where, in its trajectory, the epidemiology of HCM currently is. This is quite apart from the terminological quirk in the literature, mentioned in Section 2.2.2, under which non-fatal events have been included as endpoints and may mistakenly be counted as deaths. Fortunately the financial effect of this appears not to be large (Section 6.6).

8.2 *Range of Premium Increases*

In Section 5 we set out some baseline costs of adverse selection, and in Sections 6 and 7 we considered factors that, individually, acted to amplify or diminish these costs. Here we summarize the range of premium increases so obtained, and comment on the conditions under which the extremes might occur in reality. Recall from Section 6.1 that

premium increases of 2.5% or over in our model might be considered comparable to those in Howard (2014), although exact comparison is not possible.

The following factors amplified adverse selection losses:

(a) family history being unavailable in underwriting;
(b) a much smaller life insurance market (which may stand as a proxy for significant lapse rates in a larger market);
(c) CGT extending significantly beyond the nuclear family, approximated here by an increased average family size; and
(d) 'adverse selectors' taking out extremely large sums insured.

The following factors would tend to limit or reduce adverse selection costs:

(a) family history being available in underwriting;
(b) mutation prevalence being larger than the prevalence of clinical HCM;
(c) persons in information class ζ^{50} buying insurance at a lower rate than persons in information class ζ^{100}, when charged higher premiums because of family history;
(d) including the onset of clinical HCM as an event in the model, instead of assuming $F(20) = 1$ for early-onset mutations;
(e) unknown mutations being predominantly late-onset;
(f) epidemiology based on unselected populations becoming available;
(g) elimination of ascertainment bias from epidemiological studies; and
(h) improved treatments for HCM.

For the third of these see Haçarız (2020). The last five of these we do not model explicitly, but they should be borne in mind.

Only by combining all adverse circumstances do we get adverse selection costs of 2.5% or over. In a small market, with family history disallowed, mutation prevalence of 0.2%, annual hazard rate of fatal HCM-related events 0.0055, extensive CGT ($\lambda = 7$), 'severe' adverse selection and 'adverse selectors' taking ten times the average sum insured, our premium increases would be 8.3414%. This would appear to be far in excess of those suggested by Howard (2014). If family history is allowed the premium increases fall to 2.8159%. If we combine less extreme circumstances, for example the larger market, with family history allowed, mutation prevalence 0.9%, less extensive CGT ($\lambda = 1.8$), 'mild' adverse selection, and 'adverse selectors' taking out the average sum insured, our premium increases would be 0.0029%. This indicates the range of possibilities although ignoring that premiums based on biased mortality rates might actually be profitable.

Two particular factors merit discussion: larger sums insured; and mutation prevalence.

(a) It is clear that modelled adverse selection losses can be made arbitrarily large, by assuming that 'adverse selectors' take out large enough sums insured. The motivation for them to do so seems to be quite different from the motivation to "... secure all insurance needs before undertaking genetic testing" (Lane et al. 2015). It is not about needs, it is a monetary gamble, but a gamble with quite long odds. As we observed in Section 6.6, even with an annual mortality rate of 1%, about 2/3 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.
A 'one-shot gamble' is one that cannot be repeated, the opposite of a casino game. One way to monetize a 'one-shot gamble' would be for a large number of mutation carriers to be found, organized, and financed, by some outside agency, to buy life insurance in large amounts. The outside agency would then be on the right side of the odds, like the owner of a casino. We cannot comment on the practicalities of this, but any such scheme would seem to 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) The existing epidemiology is subject to selection and ascertainment biases that are unobservable and unknown, but are all in one direction. We would be justified in being cautious and reducing relevant biological parameters, although by how much would be arbitrary. We have not done so because adverse selection costs using any parameters that seem reasonable are small enough to make this pointless. An exception is the large difference between mutation prevalence and the prevalence of clinical HCM (Sections 7.1 and 8.1). This has been established in unselected populations and should be allowed for.

We choose as our 'headline' assumptions the following: the smaller insurance market, with family history allowed in underwriting, mutation prevalence 0.9%, annual hazard rate of fatal HCM-related events 0.0055, CGT not widespread beyond the nuclear family ($\lambda = 1.8$), and 'severe' adverse selection but with 'adverse selectors' taking out average sums insured. The resulting premium increases are 0.0382%.

8.3 *Impact on Individual Insurers*

We have modelled a life insurance market in aggregate, operating over a long period of time, during which uninsured persons can purchase insurance. Here we consider briefly how the position of an individual insurance company might be different.

(a) We assume that both population mortality and that of HCM sub-populations is fixed throughout, and moreover that insurers know accurately what they are. This allows insurers to charge 'correct' premiums, which is necessary for establishing the baseline of zero losses in the absence of adverse selection.
In reality, mortality is not fixed and insurers must update estimates from time to time, and also sometimes project future mortality.

(b) The scheme of premium payments in our model (that is, the rate of premium at any time is equal to the hazard of death) might allow an insurer to update the premiums charged to allow for future changes in mortality, especially a deteriorating experience. We have not allowed for this.

(c) An insurer with a large enough portfolio might attempt to measure emerging claims costs due to genetic disorders, and adjust premium rates accordingly. Or, some other body, such as the Continuous Mortality Investigation in the UK, might try to do so on an industry-wide basis. Indeed, by choosing premium increases as our measure of cost, we implicitly assume that insurers would attempt to do this, but we have not allowed for any such adjustments being made during the period in which our market operates.

(d) That said, most of our modelled premium increases are so small that it would be difficult or impossible reliably to detect causative effects, even in a large survival analysis. Moreover, our costs arise from the accumulation of total deaths over the entire period of the market's operation. Detecting any effects over a shorter time period would be correspondingly harder.

8.4 *Conclusions*

We have developed an explicit model of CGT in HCM and used this to explore possible adverse selection costs, expressed as percentage premium increases to recoup losses. We found the range to be very large. The very highest are comparable to, or may even be in excess of, those suggested in Howard (2014).

One key assumption in predicting the highest costs is that 'adverse selectors' desire to take out extremely high sums insured, and are not prevented from doing so by any aspect of the normal underwriting process, either medical or financial. We think this assumption is debatable. The question of whether it would be possible to monetize a 'one-shot gamble' (Thomas (2012), see Section 8.2) would be an interesting topic for research. To reach very high costs it is also necessary to make many other adverse assumptions. Perhaps the most important of these is assuming that insurers could not use family history in underwriting.

We have identified, in the published epidemiology on HCM, some features that would mitigate the possible adverse selection costs. The major one is the much higher prevalence of HCM-related mutations than of clinical HCM, suggesting lower penetrance in unselected populations. Of less financial significance, but important for life insurance applications, is the general practice among authors of HCM survival studies of using endpoints that include a substantial proportion of non-fatal events. The reason this feature is of small financial significance is the very same as the reason for doubting the attractiveness of the 'one-shot gamble'.

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. If premium rates allowing for family history are based on mortality subject to selection and ascertainment biases the costs of adverse selection may be much reduced or even eliminated.

HCM is one of several dominantly-inherited cardiomyopathies and ion-channelopathies of possible significance for life insurance. Each has an extensive published epidemiology. CGT is the form of testing used in clinical practice. Our model should be capable of estimating adverse selection costs in respect of these conditions. This is a subject for future research.

Appendices

A. Modelling Methodology

Here we outline the general modelling approach used in the past (Subramanian et al. (1999), Macdonald et al. (2003*a*), Macdonald et al. (2003*b*), Gui et al. (2006), Lu et al. (2007)) to model adverse selection in life insurance. We are unable, within the confines

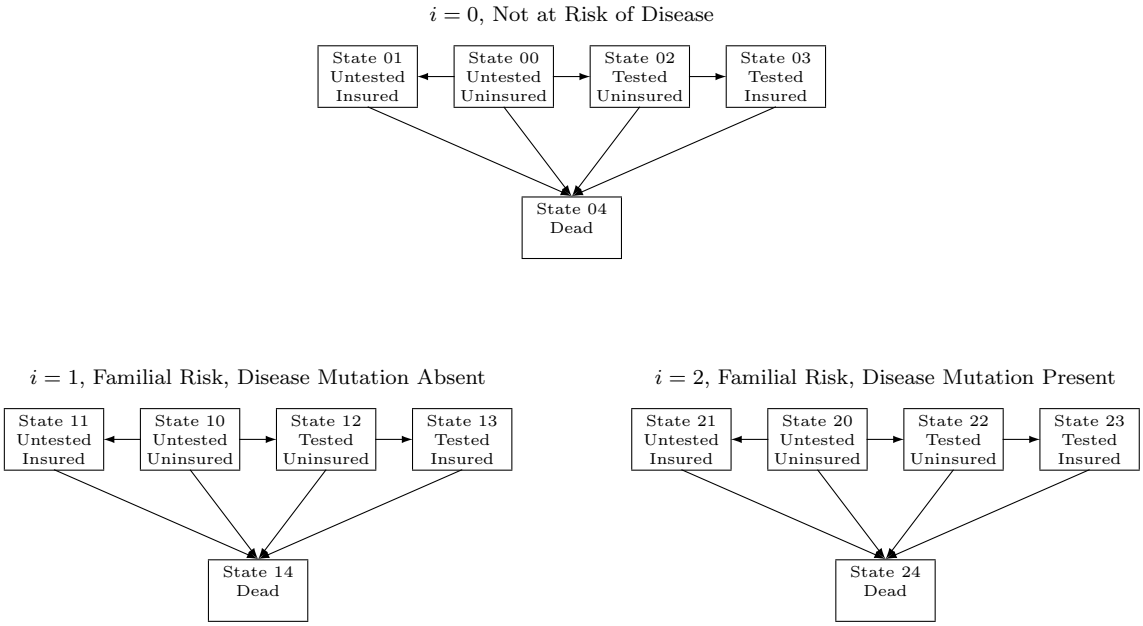

Figure 3: A mathematical model of uptake of genetic testing and purchase of life insurance for a population in which one sub-population ($i = 2$) carries a mutation associated with a dominantly inherited genetic disease.

of this paper, to supply all the details necessary to allow results to be reproduced, but the reader can find them in the papers cited above, and in the case of HCM in the PhD thesis Haçarız (2020).

The approach employs a continuous-time multiple-state model (usually but not always Markov), to represent three key features: differential mortality risk, information, and financial decisions based on that information.

Figure 3 shows a simple example, in respect of some unspecified single-gene disorder. Suppose the causative mutation has a population frequency of 1 in 1,000. Since the mutation is inherited with probability 1/2 (Mendel's laws), for every person who carries the mutation there is, on average, a sibling who does not. Thus a person chosen at random from the population is born into state 00, in sub-population $i = 0$, with probability 0.998; into state 10, in sub-population $i = 1$ with probability 0.001, and into state 20, in sub-population $i = 2$ with probability 0.001. Persons in sub-populations 1 and 2 are indistinguishable unless they have a genetic test[6]. Separating sub-population $i = 1$ from sub-population $i = 0$, though they share the same biological risks, is necessary in order to model *information* (see below).

Persons in subpopulation $i = 2$ have the higher mortality associated with the disorder; persons in sub-populations $i = 0$ and $i = 1$ do not, they have normal (population) mortality. Thus, *differential mortality* is represented.

[6]If the disorder has no cause except the mutation, then developing the disorder places a person in sub-population $i = 2$. In other cases (for example breast or colo-rectal cancer) this is not so.

Adverse selection takes place against a background of 'normal' insurance purchasing, represented by transitions from states 00 to 01, 10 to 11, or 20 to 21. A high transition intensity therefore represents a large insurance market, a low transition intensity a small insurance market. The smaller the market, the more relative impact adverse selection may have.

Suppose an event occurs, which we call W, which reveals to an individual that they must be in a family in which the mutation is inherited. For simplicity, we have omitted any such event from the model in Figure 3 but the most important example would be a first-degree relative developing the disorder, as in Section 4.4. In this case, the individual now has a family history. Suppose they have not yet bought insurance (and they would not have had a genetic test in the absence of any reason to do so). Thus, the model represents *information*, as follows.

(a) Before the event W, the individual could be in any of states 00, 10 or 20. Both the individual and the insurer would share this information.
(b) After the event W, the individual must be in one of states 10 or 20, but which is unknown. If the insurer may use family history in underwriting, both the individual and the insurer would share this information, but if not the individual has an information advantage over the insurer.
(c) If, after event W, the individual decides to have a genetic test (so moves to state 12 or 22) they know which sub-population they are in. If insurers may not use genetic test results, the individual in state 22 has an information advantage over the insurer, even if the insurer may use family history.

Finally, the model represents how *decisions* made by the individual and the insurer depend on the information they have.

(a) After event W, the insurer (if allowed) may charge a higher premium based on family history. The individual may become more likely to buy insurance (because of their greater risk) or less likely (because of the higher price). This would be represented by the transition intensities into states 11 and 21.
(b) After taking a genetic test, an individual in state 22 might be more likely to buy insurance, especially because they need not share this information with the insurer. This would be represented by the transition intensity into state 23.

Some transition intensities represent biological events (in this example, death) and others represent decisions (to be tested, to buy insurance). The former are parametrized using the best available epidemiology and a suitable population life table. The latter are not parametrized as such, from data, but are chosen to define a wide range of scenarios representing a plausible range of human behaviour.

If the genetics underlying the disorder is more complicated, for example if there are early-onset and late-onset mutations, the model in Figure 3 is easily extended by adding more pairs of sub-populations like $i = 1$ and $i = 2$ above, see Figure 2.

The main question left unanswered above, and the main source of differences between the papers cited at the start of this Appendix, and also this paper, is what the event W is and how it is to be represented in the model. In this paper, W is an HCM-related event experienced by any one of a collection of individuals forming a family (Sections 4.3 and

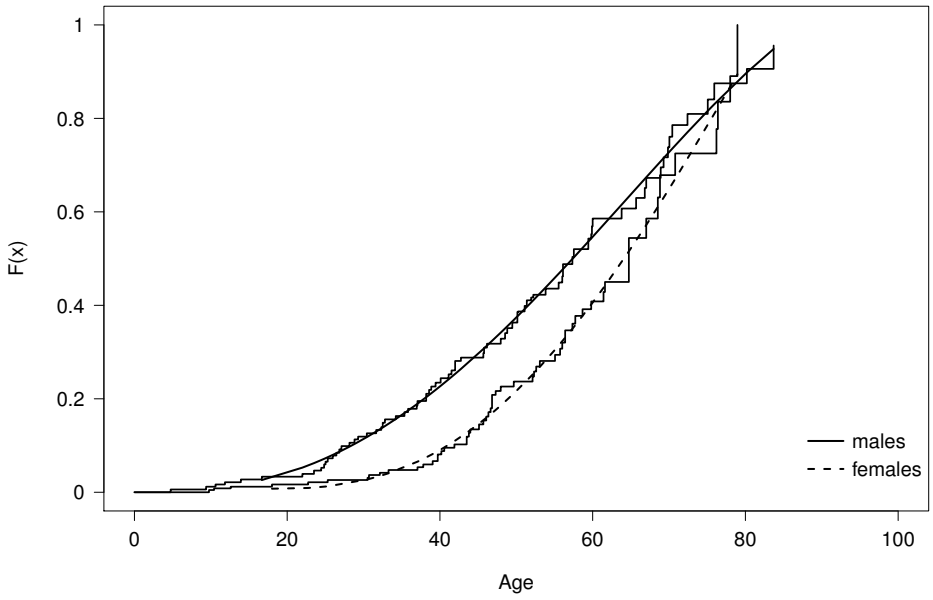

Figure 4: Late-Onset Penetrance Rate of HCM from Christiaans et al. (2011)

4.4) (compare Figure 1 with the sub-population models in Figure 3).

B. Onset Rates of HCM for Late-Onset Mutations

Late-onset HCM is mainly associated with mutations in the MYBPC3 gene (Haçarız 2020). Two studies have estimated the age-related penetrance of such mutations, Christiaans et al. (2011) and Terauchi et al. (2015). We use the first of these; sensitivity testing in Haçarız (2020) shows that the choice is unimportant.

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 4 shows their Kaplan-Meier estimates of penetrance by age x for males and females, as well as our own smoothed estimates:

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

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

The authors noted that their penetrance estimates were lower than in other studies because they excluded known clinically-affected probands and relatives.

References

Arias, E., Heron, M. & Xu, J. (2017), 'United States Life Tables, 2013', *National Vital Statistics Reports* **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.

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.

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.

Dunn, K. E., Caleshu, C., Cirino, A. L., Ho, C. Y. & Asley, E. A. (2013), 'A clinical approach to inherited hypertrophy: The use of family history in diagnosis, risk-assessment and management', *Circulation: Cardiovascular Genetics* **6**, 118–131.

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.

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.

Haçarız, O. (2020), Genetics, Insurance, and Cardiomyopathies: A Case Study of Hypertrophic Cardiomyopathy, PhD thesis, Heriot-Watt University, Edinburgh, U.K.

Hodge, S. E. (2002), Ascertainment, *in* R. Elston, J. Olson & L. Palmer, eds, 'Biostatistical Genetics and Genetic Epidemiology', John Wiley.

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).

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., Armstromg, K. & Asch, D. A. (2000), 'Pricing term insurance in the presence of a family history of breast or ovarian cancer', *North American Actuarial Journal* **4**(2), 75–87.

Lombardo, M. (2018), 'The impact of genetic testing on life insurance mortality', *The Society of Actuaries* .

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., 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. S. & 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., 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., 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 $\geq$ 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, S. M., Hellawell, J. L., Lucove, J. C., Farzaneh-Far, R. & Olivotto, I. (2016b), 'Occurrence of clinically diagnosed hypertrophic cardiomyopathy in the United States', *Americal Journal of Cardiology* **117**, 1651–1654.

Martin, J. A., Hamilton, B. E., Osterman, M. J. & Driscoll, A. K. (2018), 'Births: Final data for 2017', *National Vital Statistics Reports* **67**(8).

Office for National Statistics (2018), 'Births in England and Wales: 2017', Office for National Statistics, UK.

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.

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', *Journal of Risk and Insurance* **66**, 531–550.

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.

Acknowledgements

Oytun Haçarız is supported by a Turkish Government PhD scholarship. We acknowledge support from the Actuarial Research Centre of the Institute and Faculty of Actuaries. The authors are grateful Dr Pradip Tapadar for comments on a draft, and to the late Professor A. J. (Sandy) Raeburn and Dr Wayne Lam for many helpful discussions on the genetical and clinical aspects of HCM. We are also grateful to an anonymous referee for constructive comments that improved the paper.