THE BREAST CANCER POLYGENE AND LONGEVITY GENES: THE IMPLICATIONS FOR INSURANCE

By

Kenneth Robert McIvor

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

April 2008

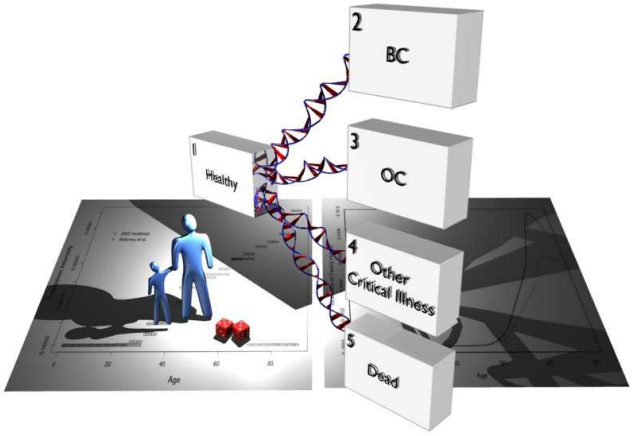
The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

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

Kenneth R. McIvor (Candidate)

Professor Angus S. Macdonald (Supervisor)

Date



For Nature, heartless, witless Nature

Will neither know nor care.

– A.E. Housman

Contents

Abstract **xv**

Acknowledgements **xvii**

Introduction **1**

1 Genetic Topics, Insurance and Numerical Tools **9**
1.1 Elementary Genetics . . . 9
1.1.1 DNA . . . 9
1.1.2 Mitochondrial DNA . . . 10
1.1.3 Genes . . . 11
1.1.4 Gametes . . . 11
1.1.5 Chromosomes . . . 12
1.1.6 Mendel's Laws . . . 12
1.1.7 The Punnet Square . . . 14
1.2 Genetic Disorders . . . 16
1.2.1 Genetic Epidemiology . . . 16
1.2.2 Single-gene Disorders . . . 17
1.2.3 Polygenic Disorders . . . 17
1.2.4 Multifactorial Disorders . . . 18
1.3 Critical Illness Insurance . . . 18
1.3.1 UK Background . . . 18
1.3.2 Coverage . . . 19
1.3.3 CI Policies . . . 20
1.4 Numerical Tools . . . 20
1.4.1 Thiele's Differential Equations . . . 20
1.4.2 Kolmogorov's Differential Equations . . . 21
1.4.3 Runge-Kutta Method . . . 22
1.4.4 Simpson's Rule . . . 22

2 The Polygenic Model and Critical Illness Insurance **24**
2.1 Introduction . . . 24
2.1.1 Breast Cancer, Ovarian Cancer and Insurance . . . 24
2.2 The Model of Antoniou *et al.* (2002) . . . 27
2.2.1 Breast Cancer and Polygenes . . . 27
2.2.2 The Hypergeometric Polygenic Model . . . 28

2.2.3 The Model of Antoniou *et al.* (2002) . . . 32
2.3 A Model for Critical Illness Insurance . . . 35
2.3.1 The Model . . . 35
2.3.2 Premiums Based on Known Genotypes . . . 36
2.3.3 An Australian Population . . . 39
2.3.4 A Comment on Genetic Tests for Polygenotypes . . . 43
2.4 Comparison of Data and Methods . . . 43
2.4.1 The Baseline Hazard . . . 43
2.4.2 Relative Risks For BRCA1/2 Mutation Carriers . . . 44
2.4.3 Penetrance . . . 44
2.4.4 Mutation Frequencies . . . 48

3 Modelling Family History with the Polygenic Model 50
3.1 Introduction . . . 50
3.1.1 Modelling Family History . . . 50
3.1.2 Definition of Family History . . . 52
3.2 Simulating Families . . . 52
3.2.1 The Simulation Model . . . 52
3.2.2 Simulating Competing Risks . . . 54
3.2.3 Sampling Insurance Applicants from Simulated Families . . . 55
3.2.4 Applicant's Genotype Distribution . . . 55
3.2.5 Premiums for an Applicant with a Family History . . . 61
3.2.6 Genotype Distributions among those without a Family History 63
3.3 Conclusions . . . 67

4 Estimating the Costs of Adverse Selection 69
4.1 Introduction . . . 69
4.1.1 The UK Moratorium on Insurers' Use of Genetic Information . 69
4.1.2 Major Genes and Polygenes . . . 70
4.2 Modelling a CI Insurance Market . . . 71
4.2.1 Model Setup . . . 71
4.2.2 A Genetic Screening Program for the Polygene Only . . . 72
4.2.3 A Genetic Screening Program for the Polygene and Major Genes 76
4.2.4 More Limited Genetic Testing for the Polygene and Major Genes 79
4.2.5 Separate Testing for Polygene and Major Genes . . . 81
4.3 Conclusions . . . 86

5 Estimating the Extent of Adverse Selection 88
5.1 Introduction . . . 88
5.1.1 A Review of Economic Modelling of Adverse Selection . . . 88
5.2 Utility Models . . . 89
5.2.1 Utility Functions . . . 89
5.2.2 Notation for the Polygenic Model . . . 92
5.3 The Purchase of Critical Illness Insurance . . . 93
5.3.1 Critical Illness Premiums . . . 93
5.3.2 Threshold Premiums . . . 94

5.3.3 Adverse Parameterisations of the Polygenic Model 98
5.3.4 Adverse Selection by Multiple Subpopulations 99
5.3.5 The Polygenotype as a Continuous Random Variable 104
5.4 Conclusions 115

6 Longevity Genes **116**
6.1 Pension Annuities and Genetics 116
6.1.1 Genes for Longevity 116
6.1.2 'Disease Genes' and Longevity 117
6.1.3 Tan *et al.* (2001) 118
6.1.4 Arking *et al.* (2005) 120
6.2 Parameter Uncertainty in the Cox Model 120
6.2.1 The Cox Model 120
6.2.2 Parameter Uncertainty in the Cox Model 121
6.2.3 A Remark on the Baseline Hazards 121
6.2.4 Sampling Distributions of Relative Risks and Premiums 122
6.2.5 Premiums for Females 123
6.2.6 Relative Risks and Premiums for Males 130
6.2.7 Relative Risks and Premiums Based on the Ashkenazi Jewish Cohort 134
6.3 The APOE Genotype and Longevity 136
6.3.1 The APOE Genotype and Mortality 136
6.3.2 Logistic Regression of Survival Data 136
6.3.3 Premium Rate Sampling Distributions Given APOE Genotype 138
6.4 Discussion and Conclusions 145
6.4.1 Acceptable Uncertainty 145
6.4.2 Acceptance Percentiles 146

7 Conclusions and Further Work **148**
7.1 Conclusions 148
7.1.1 The Polygenic Model 148
7.1.2 Longevity 150
7.2 Further Work 150
7.2.1 Realisation of the Polygenic Model 150
7.2.2 Polygenic Models in Other Diseases 151
7.2.3 Further Insurance Models 152

A Genes Conferring BC Risk **153**

B Intensities of Death and Critical Illness **158**

References **160**

List of Tables

1.1 The Punnet square for parental genotypes $AaBbCc \times AaBbCc$. The 2^3 possible gamete formations for the parents are shown along the top and down the left. 15
1.2 The matrix for parental polygenotypes $AaBbCc \times AaBbCc$ showing the genotypes' influence on cancer susceptibility. 15

2.1 The relative risks for BC and OC BRCA1 or BRCA2 mutation carriers estimated by Antoniou *et al.* (2002). The baselines are the onset rates in England and Wales in 1983–87. 34
2.2 Comparison of the incidence rates for breast cancer estimated by Antoniou *et al.* (2002) and Ford *et al.* (1998). 34
2.3 Level net premium for women, depending on polygenotype, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with the mean polygene $P = 0$. 37
2.4 Level net premium for women free of BRCA1/2 mutations, depending on polygenotype, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with the mean polygene $P = 0$. Based on an Australian population. 42
2.5 The relative risks of BC and OC for BRCA1/2 mutation carriers determined by Antoniou *et al.* (2002) and by Antoniou *et al.* (2003) in 10-year age intervals. 44
2.6 The penetrances, $q^g(x)$, for BC and OC by age 50 and 70 for BRCA1/2 mutation carriers determined by Antoniou *et al.* (2002) and Antoniou *et al.* (2003). 46
2.7 Level net premiums for CI cover as a percentage of standard risks, for BRCA1 and BRCA2 mutation carriers. Figures in brackets are the premiums from Gui *et al.* (2006) using 100% incidence rates. 47

3.1 An example of CI underwriting procedure for BC family histories. Source: Wekwete (2002) 53
3.2 Distribution of the number of daughters born in a family. Source: Macdonald, Waters & Wekwete (2003a) 54
3.3 Numbers of daughters with no family history and given major genotype, in each state in the CI model (see Figure 2.3), at selected ages. 57
3.4 Numbers of daughters with a family history and given major genotype, in each state in the CI model (see Figure 2.3), at selected ages. 58

3.5 Level net premium for females with a family history of BC or OC, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with polygenotype $P = 0$. The P + MG model uses both major gene and polygene probabilities in the weighted average EPVs, while the MG model uses only the major gene probabilities. . 61
3.6 Level net premium for females with a family history of BC or OC, as a percentage of the standard premium. The polygenic model is compared with the major-gene-only model of Gui *et al.* (2006). The latter assumed that onset rates of BC and OC among BRCA1/2 mutation carriers were either 100% or 50% of those estimated, as a rough allowance for ascertainment bias. 62

4.1 Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Screening available for the polygene only. 75
4.2 Costs of adverse selection resulting from low risk polygenotype carriers buying less insurance than normal in a critical illness insurance market open to females between ages 20–60. High risk polygenotype carriers buy insurance at normal rate. Screening available for the polygene only. 77
4.3 Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Screening available for major genes and the polygene. 78
4.4 Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Testing available for major genes and the polygene after the onset of a family history. 82
4.5 Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Separate testing for polygene and major genes. 84
4.6 Costs of modest adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Separate testing for polygene and major genes. 85

5.1 The four utility functions parameterised by Macdonald & Tapadar (2006). 90
5.2 Single premiums for various term assurances for the $P = -3$ and $P = -2$ non-BRCA mutation carrier ($M = 0$) subpopulations. 94
5.3 Premium rates X^* that are the thresholds at which adverse selection will take place, for a variety of CI policies and initial wealth $W = £100,000$. 96

5.4 Premium rates X^* that are the thresholds at which adverse selection will take place, for a variety of CI policies and initial wealth $W = £100,000$. 97
5.5 Losses at which adverse selection occurs with $\sigma_R = 1.291$, *i.e.* the $(-3, 0)$ subpopulation no longer purchase at the rate offered by the insurer. 98
5.6 Levels of σ_R at which adverse selection occurs, *i.e.* the (−3,0) subpopulation no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows. 100
5.7 Premium rates X^* that are the thresholds at which adverse selection by both the $P = -3$ and $P = -2$ polygenotype subpopulations will take place, for a variety of CI policies and initial wealth $W = £100,000$. 102
5.8 Premium rates X^* that are the thresholds at which adverse selection by both the $P = -3$ and $P = -2$ polygenotype subpopulations will take place, for a variety of CI policies and initial wealth $W = £100,000$. 103
5.9 Levels of σ_R at which adverse selection occurs within the $(-2, 0)$ subpopulation, *i.e.* the $(-3, 0)$ and $(-2, 0)$ subpopulations no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows. 105
5.10 Levels of σ_R at which adverse selection occurs when subpopulations $(-3, 0)$ and $(-2, 0)$ pool their premium, *i.e.* the (-3,0) and (-2,0) subpopulations no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows. 106
5.11 The polygenotype p^* at which adverse selection occurs for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model I utility. The figures in parentheses represent the proportion of the market who will not purchase insurance. 109
5.12 The polygenotype p^* at which adverse selection occurs for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model II utility. The figures in parentheses represent the proportion of the market who will not purchase insurance. 110
5.13 The polygenotype p^* at which adverse selection occurs under the dynamic insurer pricing method for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model I utility. The figures in parentheses represent the proportion of the market who will not purchase insurance. 113

5.14 The polygenotype p^* at which adverse selection occurs under the dynamic insurer pricing method for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model II utility. The figures in parentheses represent the proportion of the market who will not purchase insurance. 114

6.1 Genes, and their possible related disorders, that have been repeatedly studied for associations with longevity and shown significant correlations (De Benedictis *et al.*, 2001). 117
6.2 List of genes studied in Tan *et al.* (2001) labelled $g = 1, 2, \ldots, 12$; and the KLOTHO genotypes studied in Arking *et al.* (2005), labelled $g = 13, 14$. 119
6.3 The mean, standard deviation and quantiles of single premiums for a whole-life annuity for a female age 60 based on a log-normal distribution of relative risk estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$. . . . 129
6.4 The mean, standard deviation and quantiles of single premiums for a whole-life annuity for a male age 60 based on a log-normal distribution of relative risk estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$. . . . 133
6.5 The mean, standard deviation and quantiles of single premiums for a whole-life annuity for individuals age 60 with KLOTHO genotypes FF and VV based on a Normal distribution of β estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$. 135
6.6 The APOE genotypes studied in Hayden *et al.* (2005). 136
6.7 Single premiums for level whole-life pension annuities of 1 per year payable continuously, depending on APOE genotype. The premiums are expressed as a percentage of those for the most common genotype, $\epsilon 3/\epsilon 3$. Premiums are shown for healthy male and female purchasers aged 65, 70 and 75. 142
6.8 The mean, standard deviation and quantiles of single premiums for a whole-life annuity for males and females age 65 based on a log-normal distribution of relative odd estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative odds $RO = 1$. 143
6.9 Single premiums for level whole-life pension annuities of 1 per year payable continuously based on the Alzheimer's disease model of Macdonald & Pritchard (2001), treating APOE genotypes as underwriting classes. The premiums are expressed as a percentage of those for the most common genotype, $\epsilon 3/\epsilon 3$. Premiums are shown for healthy male and female purchasers aged 60, 65, 70 and 75. 144

6.10 A list of all genes/genotypes studied, and whether they are significant at a 75%, 90%, 95% or 97.5% level. A ✓ represents a significant gene/genotype and a ✗ represents a non-significant gene/genotype. The phenotype is the observable manifestation of the gene/genotype, this is either frailty or longevity. 147

A.1 List of genes which may confer additional BC risk, Rebbeck *et al.* (1999), Easton *et al.* (1999). The allele frequencies are for possible risk-conferring polymorphisms estimated from healthy Caucasian control populations and the numbers of distinct mutations are taken from the Human Gene Mutation Database. 154

List of Figures

2.1 The polygenic threshold model of Falconer (1981). Individuals whose liability is above the threshold value are affected. On average, siblings of affected individuals have higher liability than the general population. Consequently more siblings exceed the threshold value for disease. . . 29

2.2 Baseline incidence rates for BC (top) and OC (bottom) from ONS figures for England and Wales (1983–1987) and figures from Antoniou *et al.* (2002) 33

2.3 A model of the life history of a critical illness insurance policyholder, beginning in the Healthy state. Transition to the non-Healthy state d at age x is governed by an intensity $\mu^d(x)$ depending on age x or, in the case of BC and OC, $\mu^d_g(x)$ depending on genotype g as well. . . . 35

2.4 Baseline incidence rates for BC (top) and OC (bottom) from the Australian Institute of Health and Welfare (1999) and the National Breast Cancer Centre (2002), respectively 40

2.5 Baseline incidence rates for BC (top) and OC (bottom) from ONS figures for England and Wales (1983–1987) and (1973–1977) 45

3.1 The distribution of polygenotypes by major genotype among healthy daughters aged 30 and 40, with a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families. 59

3.2 The distribution of polygenotypes by major genotype among healthy daughters aged 50 and 60, with a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families. 60

3.3 The distribution of polygenotypes by major genotype among healthy daughters aged 30 and 40, who do not have a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families. 65

3.4 The distribution of polygenotypes by major genotype among healthy daughters aged 50 and 60, who do not have a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families. 66

4.1 A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing is available at an equal rate to all subpopulations. 73
4.2 Three possible behaviours of tested polygenotype carriers in the adverse selection model, labelled (a), (b) and (c). 74
4.3 A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing is available only after the appearance of a family history (FH) of BC/OC. 79
4.4 The incidence of family history for the subpopulations without BRCA mutations. A family history may not appear beyond age 50 in any subpopulation. 80
4.5 The incidence of family history for the subpopulations with BRCA1/2 mutations in the family. A family history may not appear beyond age 50 in any subpopulation. 81
4.6 A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing for major genes (MG) is available only after the appearance of a family history (FH) of BC/OC. Testing for the polygene (P) is available before a family history has appeared. 83

5.1 The four utility models given in Table 5.1 for wealth, w, between 0 and 100,000 pounds. 91
5.2 The binomial distribution with parameters (1/2, 6) (adjusted to have the mean at zero) overlaid with the Normal distribution with mean 0 and variance 3/2. 104
5.3 The Normal polygenotype distribution in the BRCA0 subpopulation. The proportions who adverse select on a 10-year term-assurance beginning at age 40 under the assumption of Model I utility are shaded in a series of overlapping greys corresponding to the loss to wealth ratio. . 107

6.1 Log-normal sampling densities of the relative risk estimates for females from Tan *et al.* (2001) for genes $g = 1, \ldots 6$ (top) and $g = 7, \ldots 12$ genes (bottom). 124
6.2 Gamma sampling densities of the relative risk estimates for females from Tan *et al.* (2001) for genes $g = 1, \ldots 6$ (top) and $g = 7, \ldots 12$ genes (bottom). 125
6.3 The empirical distributions of simulated single premiums for a whole-life annuity beginning at age 60 for female carriers. Genes $g = 1, \ldots 6$ are at the top and $g = 7, \ldots 12$ below. 126
6.4 The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 1, \ldots, 6$. 127
6.5 The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 7, \ldots, 12$. 128

6.6 The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 1, \ldots, 6$. 131
6.7 The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 7, \ldots, 12$. 132
6.8 The density curves of log-normally distributed relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for carriers of genes $g = 13, 14$. 134
6.9 The relative risk through different values of the hazard rate $\lambda_0(t)$ calculated for several relative odds values. 138
6.10 The distribution of relative risk throughout different values of the hazard rate $\lambda_0(t)$ assuming the relative odds are distributed log-normally. Graph is based on $RO_i \sim$ log-normal(0,0.25). 139
6.11 The empirical densities of whole-life annuities for a female (top) and a male (bottom) beginning at age 65, for APOE genotypes $\epsilon2/\epsilon2$, $\epsilon2/\epsilon3$, $\epsilon2/\epsilon4$, $\epsilon3/\epsilon4$, and $\epsilon4/\epsilon4$ relative to the annuity cost of a $\epsilon3/\epsilon3$ genotype carrier. 140

A.1 Forest plot of odds ratio estimates for the genes COMT, CYP17 and CYP19, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals. 155
A.2 Forest plot of odds ratio estimates for the genes CYP1A1, CYP2D6, GSTM1 and GSTP1, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals. 156
A.3 Forest plot of odds ratio estimates for the gene TP53, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals. 157

B.1 Incidence rates of other critical illnesses for males and females 159
B.2 Mortality rates, based on ELT15, with mortality after CI removed, for males and females 159

Abstract

The cost of adverse selection in the life and critical illness (CI) insurance markets, brought about by restrictions on insurers use of genetic test information, has been studied for a variety of rare single-gene disorders (adult polycystic kidney disease, colorectal cancer, Huntingtons disease and early-onset Alzheimers disease). Breast cancer (BC) has been the subject of several studies, since mutations in the BRCA1 and BRCA2 genes confer very high risk of the disease.

For the first time in any actuarial study, we consider whether the elucidation of a polygenic component of BC risk may be a crucial issue for insurers. Antoniou *et al.* (2002) fitted a polygenic model using families of BC cases. We use this model to find premium rates for critical illness insurance: (a) given knowledge of an applicant's polygenotype; and (b) given knowledge of a family history of BC or ovarian cancer. We find that the polygenic component causes large variation in premium rates even among non-mutation carriers, therefore affecting the whole population. In some cases the polygenic contribution is protective enough to reduce or remove the additional risk of a BRCA1/2 mutation, leading to cases where it will be advantageous to disclose genetic test results that are adverse in absolute terms.

We take two approaches to modelling the severity of adverse selection which may result from insurers being unable to take account of genetic tests. Firstly, we model the event history of a life, who may or may not submit to a genetic test and who may or may not purchase CI insurance, to determine what possible costs may arise for insurers given that testing is available for the polygene. Secondly, we adopt a utility model approach to infer how the genetic subpopulations may behave in regards to their insurance purchasing decision.

We also consider a number of gene variants that have been found to affect longevity. Their effects have been modelled using Cox or logistic regressions, whose fitted parameters have simple asymptotic sampling distributions. The expected present value of a life annuity allowing for these genetic risk estimates inherits a sampling distribution, which can be found by simulation. If proposing to use a genetic test as a basis to determine levels of risk, it is required that such a test should qualify as reliable and relevant. The sampling distributions of premiums give us an indication of whether this criteria is satisfied.

Acknowledgements

As a child I dreamed of becoming an architect and would promise my friends that I would one day design and build their homes. However, as I grew up I came into contact with several exceptional mathematics teachers who individually and collectively had an enormous influence on myself and my ambitions.

The most recent and most influential of these teachers was Professor Angus Macdonald. He gave me the opportunity to work with him on some very absorbing research and I have always listened carefully to the wisdom he has offered. His enthusiasm and shrewdness have repeatedly astonished me.

I am immensely grateful for the support of my friends and I apologise for my inability thus far to deliver their houses as promised. I thank those at home in Nairn and those here in Edinburgh. Psychological aid has always been at-hand from Mert and Dave. I have been lucky to have shared an office with Sing-Yee, who I have tormented over the years, and I am surprised not to have shared an office with Achilleas who has regularly kept me upbeat with his own brand of dark humour.

Thanks also deserves to go to my family. My mother, Maria, in particular has helped me relentlessly with every endeavour, and always to excess! She and the rest all know I love them.

As for Reissa, my gorgeous girlfriend, I know that very little will be beyond my reach as long as she is on my side.

I don't think I ever will get the time to build a house for each of my friends but, no matter where I am, my home will always be their house too.

Introduction

The debate surrounding genetics and insurance is of great importance to everyone. Should a pensioner's genetic profile determine their monthly income? Should the family income provider be requested to undergo a genetic test before insurance be provided? How will the decisions that we make now shape the circumstances for our children in their future? No amount of actuarial calculations can answer any of these questions outright. However, actuarial research, informed by the latest discoveries in population genetics, supplies the major policymakers with much-needed information on which to base their decisions.

There is no greater introduction to the genetics and insurance issue than that of Macdonald (2000). This paper describes how insurance markets operate by one of two basic principles: solidarity or mutuality. Solidarity is effective as an insurance principle when there is the need to maintain some basic level of insurance coverage for every individual. By this principle individuals are charged a premium which is not related to their risk but, perhaps to some other factor such as their level of income as a measure of their ability to pay. The UK National Health Service (NHS) is a prime example of the solidarity principle in action. In order to operate, the NHS requires that all individuals are obliged to have insurance, otherwise those who believe they are healthy and paying too much will opt out and eventually only the highest risks will remain insured. Alternatively, the principle of mutuality operates in a voluntary insurance market. Under the principle of mutuality those who choose to become insured band together to pool their risks for the benefit of each other. For this, each individual pays a premium that is related to the risk they bring to the pool. In order for an insurer to effect this it must obtain personal and sometimes sensitive

information from the applicant for insurance so that it may discriminate between them on the grounds of their perceived risk.

The genetics and insurance debate exists because of the ethical disputes surrounding discrimination, the maxim by which all private, voluntary insurance companies operate. Attitudes toward this have evolved over time. Early insurance companies never stratified their customers based on their smoking habit for example, but charged some basic premiums that covered individuals with a range of risk factors. However, the protocol of today's insurers is to identify the characteristics associated with higher mortality and/or morbidity risk and apportion charges appropriately. Smoking status, gender, disability and family history of disease are all factors that an insurance company may use to set their premiums. Le Grys (1997) argues that it is possible that attitudes toward genetic discrimination will alter in the same way and that it will become acceptable, but adds that there is no way to tell. Whatever standpoint society takes in the future, the situation today is that most people, do not wish to be segregated according to their genome. A MORI (Market & Opinion Research International) survey of over 1,000 people found that four out of five people believed that genetic information should not be used for setting insurance premiums (Human Genetics Commission, 2000). In opposition to this is the insurance company whose main concern is the possibility of adverse selection and the risks that stem from it.

Adverse selection, also called anti-selection, is a market process that can exist when buyers and sellers have different levels of information. In terms of insurance, adverse selection arises when, unknown to the insurer, high risk individuals (those at greater risk of death and/or disease than the general population) enter the insurance company's pool of covered risks and, after a sustained period where claims exceed actuarial expectations, force the insurer to raise the premium charged to everyone in that pool. The problem then is that the lower risk individuals in the pool will be inclined to withdraw or refuse further cover on their renewal date since the premium rate offered by the insurer is no longer fair in relation to their own risk status. Given enough amount of time, with low risk individuals filtering out and high risk individuals filtering in, eventually only the highest risks will be present on the insurer's

books. Potentially, the individuals at high risk who first pushed the premiums up may eventually find the premiums too high even for them. The insurer is at risk of making losses as it loses the advantages of bulk business and has to cover lives who have unstable mortality or morbidity risk. An insurer would argue that making losses is unprofitable to everyone, including policyholders.

Some early discussions relating to genetics and insurance in the UK addressed the need for the actuarial profession to obtain estimates of the possible costs of adverse selection. The first paper to attempt this, in respect of life insurance, was Macdonald (1997). This study gave the results of some simplistic experiments that provided bounds on the costs that may emerge from adverse selection. The most extreme cases excluded, these were given as losses of around 10% of an insurer's baseline benefits. The main conclusion from these experiments was that unlimited sums assured posed the only credible risk in the (large and mature) life insurance market. Owing to differences, such as size and benefit structure, between the life insurance market and other UK insurance markets, Macdonald (1997) stressed the futility of any attempt to extrapolate these results to other insurance products.

In response to pressure from the government, the insurers' representative body, the Association of British Insurers (ABI), announced in 1997 a code of conduct and a moratorium on the use of genetic test results for applications for life insurance of up to £100,000, if made in connection with a mortgage. Any requests for insurance beyond this limit would mean the applicant may be required to submit to genetic testing, as long as the test was deemed acceptable by the Genetics and Insurance Committee (GAIC). GAIC was established in 1998 with the task of assessing applications from the ABI (or any other body) for the use of specific genetic tests in insurance underwriting. By October 2000, the genetic test for Huntington's disease, a neuro-degenerative disorder, was deemed as sufficient by GAIC for use in assessing applications for life insurance. This test remains the only test that qualifies as reliable and relevant by GAIC to date.

In 2001 the ABI announced that the moratorium would be extended to 2006 with the new limits of £500,000 for life insurance and £300,000 for critical illness, income

protection and long term care insurance. Once again, in 2005, the moratorium was extended, this time to 2011, and the UK government, in collaboration with the ABI, published the Concordat and Moratorium on Genetics and Insurance. In the fifth report by GAIC (January 2006 to December 2006) it was reported that the ABI has said that it "may come forward with applications covering specific predictive genes for hereditary breast and ovarian cancer, but not until 2008 at the earliest". Also stated is that the ABI may be following this application with requests to extend the Huntington's disease test to critical illness and income protection insurance markets.

The first models used to determine the costs of adverse selection (Macdonald, 1997; Macdonald, 1999; Macdonald, 2000) assumed that the population could be divided into a handful of subgroups based on their genetic status. Thus each group was assumed to have a different degree of risk attributed by their genetic category. These models provided estimates of adverse selection costs, given a rough approximation of the genetic diversity in the population. The transition to research on specific genetic disorders required a greater contribution from genetic epidemiology, which, by the start of the new millennium, was growing quickly. Around 1999, the ABI had drafted a list of seven genetic disorders which they believed had the potential to harm insurance markets in the UK if testing were disallowed. Among these disorders were early-onset Alzheimer's disease, adult polycystic kidney disease, Huntington's disease, and familial breast and ovarian cancer.

Alzheimer's disease (AD) is the most common cause of dementia, however AD occurring before age 65 is rare and is known as early-onset Alzheimer's disease (EOAD). Three genes have been confirmed as causing EOAD: APP, PSEN-1 and PSEN-2. Gui & Macdonald (2002a) made estimates of the rate of onset of EOAD associated with PSEN-1 mutations, which later enabled Gui & Macdonald (2002b) to find critical illness and life insurance premium ratings given either a known mutation or a family history of EOAD, and to estimate the costs of a moratorium on genetic test results or family history information. They found critical illness premium rates to be extremely high for confirmed PSEN-1 mutation carriers, but that life insurance premiums could perhaps be offered to most known PSEN-1 mutation carriers. The effect of adverse

selection was found to be negligible except in the event of 'extreme behaviour' and a small market.

Macdonald & Pritchard (2001) looked at a major gene for AD and considered the effect that a ban on testing might have on the UK long term care insurance market. Some variants of the APOE gene put carriers at high risk of AD in their later life, leading to a greater possible need for institutionalisation. For high estimates of APOE-associated risk, their work suggested the need to 'rate-up' mutation carrier applicants by as much as 40%. They observed that the cost of adverse selection is only likely to be significant if the market is small, APOE mutation carriers are more likely to purchase insurance and genetic testing for APOE becomes widespread.

Adult polycystic kidney disease (APKD), which can lead to kidney failure and, if left untreated, death, is associated with mutations in two genes: APKD1 and APKD2. The initial actuarial study of Gutiérrez & Macdonald (2003) concentrated solely on the implications of APKD1 and APKD2 testing for critical illness insurance, and, due to data limitations (of studies pre-dating DNA-based tests), they did not differentiate between APKD1 and APKD2 mutations. Gutiérrez & Macdonald (2007) provided a more current review of the genetic epidemiology of APKD, allowing for the APKD1 and APKD2 genes, and found premium increases and costs of adverse selection in respect of both life and critical illness insurance. One of the challenges of modelling a life insurance contract which considers specific disorders for which treatment is available (in this case dialysis or kidney transplant) is that such provision is uncertain and complicates the rates of post-onset mortality. A surprising result of this study was that an individual with a family history of APKD could be expected to pay premiums greater than an individual with an adverse genetic test result for the less risky APKD2 mutation, dispelling illusions that information that is considered 'more genetic' has the greatest potential to condemn an individual's insurance application.

Huntington's disease (HD), a fatal neurological disorder, only presents in individuals who carry a faulty copy of the HD gene. It was modelled by Gutiérrez & Macdonald (2002a) whose model was applied to critical illness and life insurance models in Gutiérrez & Macdonald (2004, 2002b). The expansion of three consecutive

nucleotides in the HD gene to 36 or more repeats has been associated with earlier age-at onset of disease, so this was the first actuarial study to consider insurance pricing in the presence of a variable age-at-onset mutation. The authors found that individuals with a minimally-expanded mutation (36–39 repeats) may be able to obtain insurance (life and critical illness) at standard rates. They cautioned that such a situation could cause problems in what has been termed a 'lenient' moratorium, where individuals with a family history and tested 'clear' may obtain insurance at standard rates. By offering a reduced rate not just to those tested 'clear', but also those tested and found to be low risk mutation carriers, effectively removing more lives from the risk pool of those with a family history, the premiums charged to those with a family history would be expected to rise. But should this happen even further, say by offering reduced rates to medium risk mutation carriers, the insurer is, in effect, discriminating on adverse genetic test results, and permitted discrimination has upset the intentions of the moratorium.

The development of breast cancer (BC) or ovarian cancer (OC) can be classified as either sporadic or hereditary. Hereditary BC and OC is associated with two genes called BRCA1 and BRCA2. These diseases have attracted more actuarial studies than any other genetic disorder. Studies include Gui *et al.* (2006), Macdonald, Waters & Wekwete (2003a, 2003b), Lemaire *et al.* (2000) and Subramanian *et al.* (1999). It is the consensus among these studies that positive genetic tests for either of the two BRCA genes would require raised premiums, or even declinature, for life or critical illness insurance. On the other hand, a family history of BC or OC does not warrant such severe measures as do other genetic disorders since BC and OC are not caused entirely by genetic factors, so a history of non-hereditary BC or OC may develop by chance. For instance, only about 2% of all BC cases are associated with BRCA1 and BRCA2 mutations. The works of Lemaire *et al.* (2000) and Subramanian *et al.* (1999) were undertaken before very specific epidemiological data were available for the BRCA1 and BRCA2 genes so these studies concentrated on the insurability of individuals with family histories, or with unspecified BRCA mutations. Macdonald, Waters & Wekwete (2003b) were able to calculate critical illness premiums for those

carrying BRCA1 or BRCA2 mutations, and for those who have a family history. Gui *et al.* (2006) was the first study to consider the development of a BC/OC family history as an event in an individual's life with an associated intensity of onset, and apply this to pricing life and critical illness policies.

The work in this thesis is based primarily on a new genetic model of BC and OC: the polygenic model. We consider the implications of this model for critical illness insurance.

The first chapter is a technical introduction to the fundamentals of basic genetics, the UK critical illness insurance market, and the methods that are employed throughout the thesis.

In Chapter 2 the polygenic model is defined. We build a critical illness insurance pricing model based on UK-specific intensities of BC, OC, other critical illnesses and mortality. We use this pricing model in conjunction with the results of a fitted polygenic model supplied by Antoniou *et al.* (2002) to compute the premiums for a critical illness policy offered to carriers of each of the modelled genotypes.

UK insurers are allowed to use family history information to underwrite applicants for critical illness insurance. In Chapter 3 we simulate the lifetimes of individuals in large numbers of families to approximate the frequencies of genotypes in the population of individuals applying for critical illness insurance. It is possible to compare the genotype frequencies of individuals with and without a family history. This allows us to find the premiums that should be charged to those with a family history.

In Chapter 4 we investigate the implications of a moratorium on using adverse genetic test results. To do this we set up a model of a UK critical illness insurance market and find the proportion by which all premiums must rise in order to negate the extra costs created by those who adverse select. We consider random testing in the general population and, by including the incidence rates of developing a family history (calculated in Chapter 3), testing which is only offered to those with a history of BC or OC in their family. We consider an intermediate of these cases and also make assumptions on the behaviour of tested individuals ranging from modest to extreme.

By assuming that the population's desire to insure can be modelled by utility

functions, in Chapter 5 we map out some of the circumstances (levels of possible losses, severity of the polygenic model, *etc.*) which will result in adverse selection. The same framework allows us to calculate the proportion of the population who will refuse to purchase insurance as a result of high risks entering the pool. This is done in a setting where premiums are set and fixed indefinitely by the insurer and in a setting where premiums vary in accordance with the critical illness risk of individuals who are still prepared to obtain cover.

The penultimate chapter, Chapter 6, deals with the genetics of longevity and the impact that this may have in a pensions market. The central focus here is on the reliability of estimates of the risk conferred by a gene, based on small-scale genetic studies. This chapter uses the sample relative risk estimates of some studies, which fit the Cox model to lives tested for an assortment of genes with suspected involvement in the determination of lifespan, to find the corresponding sampling distribution of whole-life annuity prices. Similar calculations are made using the sample odds ratio estimates from a logistic model.

A shared characteristic between the genes that influence an individual's longevity and the genes that are part of the polygenic model is that both types confer only modest risk in isolation. Our belief is that such genes may have serious implications for insurance when considered altogether. However, although we find longevity to be modified by several genes, we lack the epidemiological evidence to know if the combined effect of the genes is a serious risk to insurers. Another similarity is that most geneticists believe there to be a common biology between cancer and longevity (Finkel, Serrano & Blasco, 2007). Much of this belief is based on several theories regarding the genetic mechanisms that underlie both cancer and ageing.

Conclusions and suggestions for further work are given in Chapter 7.

Chapter 1

Genetic Topics, Insurance and Numerical Tools

1.1 Elementary Genetics

1.1.1 DNA

DNA, or Deoxyribonucleic acid, is a molecule found in nearly all living creatures. It determines the form and function of the cell and carries genetic information forward into the next generation of offspring. It can be found in the nucleus of a eukaryote cell and in the cytoplasm of a prokaryote cell.

DNA is composed of two complimentary chains of nucleotides which, when joined in sequence, produce a connected double helix. Each nucleotide has a deoxyribose-phosphate link, the outer section of the double helix, and one of four 'bases', the strands that unite the helices. The bases are organic compounds which can be either adenine, cytosine, guanine, or thymine, denoted A, C, G, or T, respectively. The human genome, consisting of all DNA within a single cell nucleus, contains approximately 3.2 billion base pairs. Because of their shape, bases may only bond A to T and C to G. However, along the helix, bases may lie in any order, and that order is important since it constitutes the code that produces life with all its varieties.

The nucleotides on the DNA strand code for proteins which are all constructed

from an array of only 20 amino acids. When DNA is required to produce a protein product the code is read by splitting the nucleotides into groups of three. Each group is referred to as a codon and contains either code for an amino acid or code denoting the end of a coding region (known as a stop or termination codon). The codons are read by polymerase enzymes which synthesise the ribonucleic acid (RNA) used to transfer amino acids to the ribosome; this is known as transcription. The ribosome provides the structural support for the protein product (or polypeptide chain). With the correct composition of amino acids, the cell obtains the required protein and the process of protein production is complete.

1.1.2 Mitochondrial DNA

In Section 1.1.1 DNA contained exclusively within the cell nucleus, called nuclear DNA, was described. However there are two types of DNA carried by humans. The other type is called mitochondrial DNA, which is contained within the mitochondria of the cell. The mitochondrion is a membrane-enclosed organelle that is positioned outwith the nucleus and which generates a source of chemical energy for the cell. The mitochondrial genome differs from the nuclear genome quite significantly; it consists of only 16,569 base pairs and is normally inherited exclusively from the mother. This makes mitochondrial DNA a powerful tool in tracing maternal lineage. In endosymbiotic theory, it is believed that mitochondria originated outside of humans and that at some point the human cell assimilated mitrochondria (which has a bacteria-like structure) and the two were able to exist successfully in a symbiotic relationship.

Mitochondrial DNA is of special interest since it is believed to be of great importance to the study of longevity (Santoro *et al.*, 2006). This relationship is thought to be primarily due to the susceptibility of mitochondria to oxidative damage, a consequence of cell metabolism. It is believed that there exist forms of mitochondrial DNA that offer greater resistance to this damage than others.

1.1.3 Genes

If we think of DNA as the letters and words of the genetic code, then genes are the instruction manuals for a gene product which are written in that code. There are approximately 20,000 – 25,000 genes in the human genome.

The location of a gene is referred to as a locus and at each locus there are two copies of the gene. If these copies are identical they are called homozygous and if they are non-identical they are called heterozygous. Copies of these genes can differ because of mutation. The different versions of the gene caused by mutation are called alleles or gene variants. Genes with only two possible types are called bi-allelic, those with more are called multi-allelic.

Mutations are rare (less than 1% population frequency) random errors in the base pair sequence. A change which is more common than this is often termed a polymorphism. There are two forms of mutation or polymorphism: germline, one that is passed through the sex cells to descendents, or somatic, one that occurs in the non-sex cells and is not heritable. A mutation may be a point mutation, which replaces a nucleotide base, an insertion, which adds a nucleotide, or a deletion, which removes a nucleotide.

The composition of alleles at a locus or set of loci is known as the genotype. Through complex biological processes and interaction with the environment the genotype may manifest itself in what is known as the phenotype, an observable trait or disease.

It is common for a gene to be referred to by the disease it is associated with. For example, when mutated the gene IT15 is a high risk factor for Huntington's disease and hence is better known as the HD gene. Despite the label "disease gene" it should be borne in mind that all of us carry the HD gene, except in the majority it is not mutated and is operating successfully.

1.1.4 Gametes

The transmission of genes from parent to child is conducted via germ cells known as gametes. Gametes are created during a cell division process known as meiosis. The

female produces the larger gamete called the ovum (or egg) and the male produces the smaller gamete called a spermatozoon (or sperm cell). These gametes are haploid cells which means they only contain one complete set of chromosomes whereas the human diploid cell contains two. Under independent assortment (see Section 1.1.6) the genes contained on the gamete chromosomes are independent random selections from the original cell. When the gametes successfully unite they form the diploid zygote which contains one set of the female's chromosomes and one set of the male's. The zygote eventually develops into an embryo. This is how the offspring inherits parental genes.

1.1.5 Chromosomes

Large lengths of continuous DNA are packaged in chromosomes. There are 23 paired sets of chromosomes in each cell. The first 22 pairs are called autosomal and the last is the sex chromosome. In a female the sex chromosome pair is composed of two copies of the same chromosome, the 'X' chromosome. A male on the other hand carries one 'X' and one 'Y' chromosome. Genes can be categorised as autosomal, X-linked, or Y-linked, depending on their chromosomal location.

Some genetic disorders do not derive from code mutations but from chromosome abnormalities. These fall into two categories: numerical or structural abnormalities. A numerical abnormality is when the number of copies of an individual chromosome is more or less than the standard of two. For example, if an individual carries three copies of chromosome 21 (or trisomy 21), that individual will be born with Down Syndrome. A structural abnormality relates to errors in segments of individual chromosomes such as deletions, duplications, translocations, inversions and the formation of rings. Most chromosome abnormalities are not in fact inherited but occur as an accident in the egg or sperm.

1.1.6 Mendel's Laws

Gregor Johann Mendel is commonly revered as the "father of modern genetics" due to his pioneering research into the inheritance of pea plant traits in the mid-19th century.

In 1866 Mendel published his paper, *Experiments in Plant Hybridisation*, which did little to impress the scientific community of the time and received much criticism. It was not until 1900, several years after his death in 1884, that the significance of his work was recognised, prompting the foundation of a new science known now as genetics.

From his work with the pea plants Mendel was able to establish a set of basic tenets relating to the transmission of traits. These can be divided into two laws of inheritance:

The Law of Segregation

Also known as Mendel's First Law, the Law of Segregation describes the way that genes separate from parental genomes and combine to produce that of the child's. The law is composed of four parts:

(a) Genes may take alternative forms, known as alleles.

(b) Each organism inherits two alleles, one from each parent.

(c) If the two alleles are different, one will be dominant for the trait and the other recessive.

(d) When gametes are produced the pairs of alleles separate so that the gamete only contains a single allele.

The Law of Independent Assortment

The Law of Independent Assortment states that when allele pairs separate to form gametes they do so independently. This implies that heritable characteristics are transmitted independently. However, due to a phenomenon known as linkage, this law does not necessarily hold true for genes on the same chromosome.

Mendel learnt that some traits in his pea plants could be dominant or recessive. To be dominant means that a trait or disease may manifest in someone heterozygous

for the associated gene. A recessive trait or disorder will only appear when the associated gene is homozygous. There are intermediates of these properties known as codominance and semi-dominance.

Much of the early work in genetics was concerned with the transmission of personal traits such as height and eye colour, and as a result the language is often geared towards this. It requires only a small step in innovation to extend this to thinking about the manifestation of a disease. A key distinction to make when discussing the genetics of disease is in the classification of genes as either *deterministic* genes or *susceptibility* genes. A carrier of a deterministic gene would most likely develop the disorder in their lifetime, and we say that the gene is *fully penetrant*. A carrier of a susceptibility gene on the other hand may never exhibit symptoms of the disease but would be at higher risk than those without the gene. The gene in this case is said to have *incomplete penetrance*.

1.1.7 The Punnet Square

A Punnet square is a matrix used to represent all possible combinations of parental alleles and find the frequency of each configuration. The following is an example of how the Punnet square might be used to describe the transmission of several parental genes to a child.

We assume that there are three loci, locus 1, locus 2 and locus 3, and at each locus there exist two genes, one inheritied from the mother and one from the father. We assume that each of these genes is bi-allelic and therefore can only take one of two forms: mutated or non-mutated. For locus 1 let us denote a mutated gene as having allele A and a non-mutated gene as having allele a, and likewise for locus 2 and 3 alleles B, b, C and c, respectively.

Suppose that each allele is equally common: that is, a randomly chosen individual in the population has the mutated allele with probability 1/2 and the non-mutated allele with probability 1/2. These variants then are in fact polymorphisms. By considering the mating of two parents who both have heterozygous genotypes $AaBbCc$ we can see all the possible combinations that could be seen in the offspring. This can

be shown by using the Punnet square. Table 1.1 illustrates the Punnet square for our example.

Table 1.1: The Punnet square for parental genotypes $AaBbCc \times AaBbCc$. The 2^3 possible gamete formations for the parents are shown along the top and down the left.

	ABC	*ABc*	*Abc*	*aBC*	*abC*	*aBc*	*AbC*	*abc*
ABC	*AABBCC*	*AABBCc*	*AABbCc*	*AaBBCC*	*AaBbCC*	*AaBBCc*	*AABbCC*	*AaBbCc*
ABc	*AABBCc*	*AABBcc*	*AABbcc*	*AaBBCc*	*AaBbCc*	*AaBBcc*	*AABbCc*	*AaBbcc*
Abc	*AABbCc*	*AABbcc*	*AAbbcc*	*AaBbCc*	*AabbCc*	*AaBbcc*	*AAbbCc*	*Aabbcc*
aBC	*AaBBCC*	*AaBBCc*	*AaBbCc*	*aaBBCC*	*aaBbCC*	*aaBBCc*	*AaBbCC*	*aaBbCc*
abC	*AaBbCC*	*AaBbCc*	*AabbCc*	*aaBbCC*	*aabbCC*	*aaBbCc*	*AabbCC*	*aabbCc*
aBc	*AaBBCc*	*AaBBcc*	*AaBbcc*	*aaBBCc*	*aaBbCc*	*aaBBcc*	*aABbCc*	*aaBbcc*
AbC	*AABbCC*	*AABbCc*	*AAbbCc*	*AaBbCC*	*AabbCC*	*AaBbCc*	*AAbbCC*	*AabbCc*
abc	*AaBbCc*	*AaBbcc*	*Aabbcc*	*aaBbCc*	*aabbCc*	*aaBbcc*	*AabbCc*	*aabbcc*

We now want to think of the non-mutated allele a as cancer protecting and the mutated allele A as cancer predisposing, and we use the same lower and upper case notation to represent the cancer risk of the alleles at the other two loci. Since the genotype that confers neither excess nor reduced risk is one with three cancer predisposing and three cancer protecting alleles (*e.g.* $AaBbCc$) we assign this a numerical genotype value of 0. Any additional cancer predisposing alleles to this 'standard' genotype increases the numerical value by one and any additional cancer protecting alleles reduces the numerical value by one. This implies that an individual who holds two copies of the A allele, two copies of the B allele and two copies of the C allele has a genotype that confers the maximum cancer risk, in this case a genotype with numerical value 3. We can transform all the possible offspring polygenotypes shown in Table 1.1 to their equivalent numerical genotypes. These are given in Table 1.2.

Table 1.2: The matrix for parental polygenotypes $AaBbCc \times AaBbCc$ showing the genotypes' influence on cancer susceptibility.

	ABC	*ABc*	*Abc*	*aBC*	*abC*	*aBc*	*AbC*	*abc*
ABC	3	2	1	2	1	1	2	0
ABc	2	1	0	1	0	0	1	-1
Abc	1	0	-1	0	-1	-1	0	-2
aBC	2	1	0	1	0	0	1	-1
abC	1	0	-1	0	-1	-1	0	-2
aBc	1	0	-1	0	-1	-1	0	-2
AbC	2	1	0	1	0	0	1	-1
abc	0	-1	-2	-1	-2	-2	-1	-3

There are 2^3 different gamete formations possible from each parent and this gives us $2^3 \times 2^3 = 64$ cells in the Punnet square. By counting the occurrences of each genotype in Table 1.2, the ratio of genotypes $-3, -2, -1, 0, 1, 2, 3$ in the offspring is 1:6:15:20:15:6:1. Alternatively we could say that the frequency of the genotypes follows a binomial distribution with paramenters $n = 6$ and $q = 1/2$.

We will return to this topic in Section 2.2.2.

1.2 Genetic Disorders

1.2.1 Genetic Epidemiology

Genetic epidemiology seeks to determine the rôle of genetics and environmental factors in the development of a particular trait or disease. The science is relatively young; genetic epidemiology surfaced in the 1960s with the union of statistical genetics and classical epidemiology. The progression of discoveries in the science is highly correlated with the pace of technological advances (for example, in DNA sequencing) in which there has been much progress in the past ten years.

Genetic epidemiologists study disorders by gathering groups of related individuals, called pedigrees in the genetics literature, and collect data on their DNA, lifestyle and environment. To obtain a pedigree the epidemiologist must begin by recruiting a single individual, who is called the proband, and then attempt to bring as many of their family members as possible into the study. The simplest way of gathering these pedigrees would be to choose probands randomly from the population. However, most genetic disorders are rare, so to sample randomly would harvest very few affected individuals (the individuals who will provide the most information about the disease and associated genes) and would be a waste of resources. The alternative option then is to select families that already contain affected members. This means that the epidemiologist has access to a high risk population in which the disease-causing allele (if such an allele exists) will be more common than in the general population.

The problem that arises when the epidemiologist ascertains only affected probands is that we do not get the full picture of the disease in the population. The epidemi-

ologist would like to recruit a representative sample of all individuals who carry the susceptibility gene but, by selecting only affected individuals, those that carry the gene and are yet to become affected are missed. The gene conferring risk of the disease will seem to have greater effect than it actually does and it will seem to be more common in the population then it actually is. This problem is called ascertainment bias (see Burton *et al.*, 2001 for further details) and several methods have been proposed to deal with it (Hodge & Vieland, 1996; Rabinowitz, 1996; Cannings, 1977; Elston, 1973).

In this section we shall describe three categories of genetic disorders: single-gene, polygenic and multifactorial.

1.2.2 Single-gene Disorders

Single-gene, or monogenic, disorders are caused by a defect in one gene alone. These disorders are individually very rare but in total affect about 1% of the population.

It is often easy to predict the risk of inheriting the gene associated with a single-gene disorder since it will follow a Mendelian pattern of inheritance and exhibit the properties outlined in Section 1.1.6. However, not all single-gene diseases are fully penetrant and so do not always have an observable phenotype.

Common examples of single-gene disorders are Huntington's disease (autosomal dominant), cystic fibrosis (autosomal recessive) and muscular dystrophy (X-linked recessive).

1.2.3 Polygenic Disorders

A polygenic disorder is one that derives from the contribution of several genes that have common variants in the population. The example we used in Section 1.1 relates to a polygenic condition. The term 'polygene' encapsulates all genes which are believed to be responsible for the polygenic disorder.

The scale of investigation into polygenic disorders has escalated since the sequencing of the human genome, which made available the range of genetic variation across many loci. However, it is often difficult to identify the genes that constitute the poly-

gene for a given disease since individually they only confer a small amount of risk for the disease.

Although polygenic disorders do tend to "run in families", due to complex interactions between the genes in the polygene (known as epistasis), the pattern of inheritance does not conform to the basic rules of Mendelian inheritance for qualitative traits.

Cancer is a good example of a polygenic disorder. It is believed that the risk of onset of many cancers is related to the status of several genes along the genome.

1.2.4 Multifactorial Disorders

Very few disorders are purely Mendelian, purely polygenic or purely environmental. Most have a mix of factors which contribute towards risk of onset and we call these multifactorial disorders. The best known example of a multifactorial disorder is heart disease which is a product of several lifestyle factors (diet, exercise, smoking status, *etc.*) and several genes. Most common diseases can be classed as multifactorial.

The UK Biobank project (www.ukbiobank.ac.uk) is a medical research initiative which aims to discover the mechanisms behind many multifactorial diseases. UK Biobank will recruit 500,000 individuals aged 40 to 69 and follow them up for 10 years. Genetic samples will be taken from each volunteer so that the association between genes, environment and disease may be better understood.

1.3 Critical Illness Insurance

1.3.1 UK Background

Traditionally, Critical Illness (CI) insurance is a contract between an individual and an insurance company whereby the insurer agrees to pay a lump-sum to the individual, on diagnosis of a qualifying disease within the term of the policy. This agreement is made in exchange for regular payments made up to either disease onset or end of the policy term, depending on which occurs first. The benefit is often used to pay for a medical procedure but the ultimate decision of how it is spent lies in the hands of

the insured. Over time the product has evolved and now policies can be found which pay benefits as regular payments with or without a lump-sum or pay benefits based upon the performance of a surgical procedure. Coverage is available to policyholders as individuals or as a group.

The UK CI market is very new. Lloyds Life issued the first product in 1985 with little success. A year later, however, the CI product began to profit and expand when sold as a rider to a whole life policy. By the 1990s over 50 UK companies were offering the product. Annual policy sales peaked at over one million in 2003 but stood at around half that in 2005 after concern about premium uncertainty and slow-down in the mortgage market (Hannover Life Re (UK), 2006). In 2005, CI insurance was attracting as much as 16% of the premiums that the life insurance market attracted (ABI, 2005).

1.3.2 Coverage

A number of illnesses are covered in a CI policy, but this varies greatly from company to company as each attempts to differentiate their product. Ideally three criteria should be satisfied for a disease to be an appropriate inclusion in CI cover. It should:

(a) be perceived by the public as being both serious and common,

(b) have a commonly agreed and unambiguous definition,

(c) be accompanied by sufficient data on which to price the policy appropriately.

Some of the disorders that satisfy these criteria are cancer, heart attack, stroke, coronary artery bypass graft, multiple sclerosis and kidney failure.

At the top of the list of most common CI claims is cancer. In 2006, CI claims were composed of 59% cancer, 15% heart attack, 8% heart surgery related and 7% from stroke (Skandia UK, 2007). While cancer made up 48% of males' CI claims, 77% of females' claims were due to cancer. The number of claims made after onset of breast cancer in females amounted to 58% of all females' claims.

1.3.3 CI Policies

There are two main categories of CI policy, namely stand-alone and accelerated benefits:

(a) A stand-alone CI plan only provides cover against CI. A death benefit is not normally paid although some plans may pay a nominal amount or return premiums.

(b) An accelerated benefit CI plan is essentially a life insurance policy which will pay the sum assured early if a CI occurs. Although most policies pay the entire sum assured on the event of a CI not all are accelerated like this.

Past UK sales experience has proved that accelerated benefit plans are more popular than stand-alone policies.

It is also possible to obtain an extension to the accelerated benefit plan that allows the buyer to obtain a reinstatement of the death benefit after survival of a CI. This is known as a buy-back option.

1.4 Numerical Tools

1.4.1 Thiele's Differential Equations

Norberg (1995) formulated a differential equation for the moments of present values of payments in a continuous time Markov chain. For a life in state j, μ_t^{jk} is the transition intensity of moving to state k (with $\mu_t^{j\cdot} = \sum_{k\neq j} \mu_t^{jk}$), δ_t^j is the force of interest, b_t^j is the payment payable continuously while within the state and b_t^{jk} the payment payable on transition to state k. A negative value of b_t^j represents a premium being paid. The differential equation for the qth moment of the present value of the payments $V_t^{(q)j}$ in state j is given by:

$$\frac{d}{dt}V_t^{(q)j} = (q\delta_t^j + \mu_t^{j\cdot})V_t^{(q)j} - qb_t^j V_t^{(q-1)j} - \sum_{k\neq j}\mu_t^{jk}\sum_{r=0}^{q}\binom{q}{r}(b_t^{jk})^r V_t^{(q-r)k}. \qquad (1.1)$$

Most often we are only interested in the expected value of the payments (so that we can find the premiums payable using the Equivalence Principle). The first moment ($q = 1$) is better known as Thiele's differential equation (Hoem, 1988):

$$\frac{d}{dt}V_t^j = \delta_t V_t^j - b_t^j - \sum_{k \neq j}(b_t^{jk} + V_t^k - V_t^j)\mu_t^{jk}. \tag{1.2}$$

Since no reserve should be held at the end of the policy (time T), the boundary condition is:

$$V_T = 0. \tag{1.3}$$

1.4.2 Kolmogorov's Differential Equations

In a simple two-state Markov model with transition possible in only one direction (alive, a, to dead, d), with intensity μ_t^{ad} we obtain the differential equation for the probability of survival from time x to $x + t$, given survival up to time x:

$$\frac{d}{dt}\ {}_tp_x^{aa} = -{}_tp_x^{aa}\mu_{x+t}^{ad}, \tag{1.4}$$

which when solved gives:

$${}_tp_x^{aa} = \exp\left(-\int_0^t \mu_{x+s}^{ad} ds\right), \qquad {}_tp_x^{ad} = 1 - \exp\left(-\int_0^t \mu_{x+s}^{ad} ds\right). \tag{1.5}$$

However, modelling more sophisticated state-spaces requires a more general method.

Kolmogorov's differential equation provides a generalised formula for the relationship between the transition intensities and the survival probabilities. For a general multiple-state Markov model, ${}_tp_x^{ik}$ is found by solving:

$$\frac{d}{dt}\ {}_tp_x^{ik} = \sum_{j \neq k}\left({}_tp_x^{ij}\mu_{x+t}^{jk} -{}_t\, p_x^{ik}\mu_{x+t}^{kj}\right) \qquad \text{for } i \neq k. \tag{1.6}$$

The boundary conditions are:

$${}_0p_x^{ik} = \delta_{ik}, \tag{1.7}$$

where δ_{ik} is the Kronecker delta.

1.4.3 Runge-Kutta Method

As a numerical method to approximate a solution to Thiele's and Kolmogorov's differential equations we employ the fourth-order Runge-Kutta method. This is used in place of Euler's method as it is optimal in terms of error per step and numerical stability.

For the differential equation $y' = f(t, y)$, with initial condition $y(0) = y_0$ and stepsize h, the fourth-order Runge-Kutta method has iterations of the form:

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f(t_n + h/2, y_n + hk_1/2), \\ k_3 &= f(t_n + h/2, y_n + hk_2/2), \\ k_4 &= f(t_n + h, y_n + hk_3), \end{aligned}$$

$$y_{n+1} \approx y_n + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right), \tag{1.8}$$

which are repeated to approximate y. For further details see Press *et al.* (2002).

Although it is possible to implement an adaptive step-size routine, we obtain sufficient accuracy by fixing the step-size $h = 0.0005$. Since the boundary condition (Equation (1.3)) for Thiele's differential equation is at the end of the policy term, we must run the iteration backwards to obtain the solution.

1.4.4 Simpson's Rule

Throughout this work we will meet functions where the antiderivative cannot be written in elementary form. One of these functions is the probability density function of the normal distribution, where the cumulative probability distribution is not available explicitly. To obtain approximations of the integrals we adopt a numerical integration method.

Two of the most popular methods available for numerical integration are the Trapezoidal rule and Simpson's rule. The Trapezoidal Rule uses straight lines to follow the curvature of the function and then calculates the area below this (the sum of the areas of a series of trapezoids). This would be an ideal method for integrating a linear function, however not so for higher dimensional polynomials where the straight lines provide a poor fit. A more efficient method would be to approximate the curvature of the graph by a parabola. This is achieved using Simpson's rule.

The integral $\int_a^b f(x)dx$ is approximated by breaking the interval $[a, b]$ into $2n$ pieces (divided by the points $x_0, x_1, \ldots, x_{2n}$) of width $h = (b-a)/2n$ and fitting a quadratic between each consecutive group of three points. The result is the approximation:

$$\int_a^b f(x)dx \approx \frac{h}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{2n-1}) + f(x_{2n})\right). \tag{1.9}$$

See Burden & Faires (1997) for further details.

Chapter 2

The Polygenic Model and Critical Illness Insurance

2.1 Introduction

2.1.1 Breast Cancer, Ovarian Cancer and Insurance

Breast cancer (BC) is the most common cancer among women in the UK; one in nine women develop BC in their lifetime. Ovarian cancer (OC) is the fourth most common cancer among women, and the UK has the highest incidence of OC in Europe (Cancer Research UK). Together they account for a significant proportion of claims under critical illness (CI) insurance policies (see Section 1.3.2). It is well known that mutations in either the BRCA1 or BRCA2 genes can increase the risk of BC or OC at early ages very substantially.

The genetic risk associated with family histories of BC or OC has prompted more actuarial research than has any other genetic disorder. The work has built upon the genetic epidemiology of BC and OC, which is still developing. Early epidemiological studies selected highly affected families; these were the basis for actuarial studies by Subramanian *et al.* (1999), Lemaire *et al.* (2000), and Macdonald, Waters & Wekwete (2003a, 2003b). Recent advances in the epidemiology include larger sample sizes and less biased selection of subjects or families. A recent actuarial study allowing for these

is Gui *et al.* (2006). The aim of all these actuarial studies has been to model how life and CI insurance pricing may be affected: (a) if the insurer knows of the genetic risk; or (b) if the applicant for insurance knows of the genetic risk but the insurer does not.

In the UK, the Genetics and Insurance Committee (GAIC) has the task of assessing applications made by the insurance industry to be allowed to use genetic test results in underwriting, provided: (a) the test results were known because of past clinical history; and (b) the sum assured exceeds the limit set in an agreed moratorium (currently £500,000 for life insurance and £300,000 for CI insurance). Because of their significance, tests for BRCA1/2 mutations are very likely to be the subjects of applications to GAIC.

UK insurers are still allowed to use family history in underwriting (unlike in some other countries, such as Sweden) so in view of the high limits set by the moratorium, the vast majority of applications involving a family history of BC or OC will continue to be underwritten on that basis. Although genetic test results have attracted much attention, the implications of a family history are of more practical importance.

The main epidemiological quantity needed for actuarial modelling is the rate of onset, here denoted $\mu_g(x)$. This is the force of onset of the disease (or hazard rate) at age x, for a person with genotype g. If estimates of $\mu_g(x)$ are available, they can be incorporated in a multiple decrement model for CI insurance almost trivially, or more generally given any payment function we can compute its expected present value (EPV), denoted $X(g)$. However, this assumes the genotype g to be known. If all that is known is the existence of a family history when a woman age x applied for insurance, the corresponding EPV, denoted $X(f)$, is:

$$X(f) = \sum_g \mathrm{P}[\text{ Genotype is } g \mid \text{Family history exists at age } x \text{ }]\, X(g) \qquad (2.1)$$

where the sum is over all possible genotypes g. Thus the genotype-specific quantities are still needed, even if the focus is on family history. An important point, which

will drive our choice of methodology later in Chapter 3, is that the conditional genotype probabilities in Equation (2.1) usually depend on the transmission probabilities, namely the probabilities that a child of parents whose genotypes are known will have any given genotype.

Another key feature of the earlier genetic epidemiology of BC and OC is that it was based upon the inheritance of major genes, namely single genes in which mutations are alone sufficient to cause the disease. BRCA1 and BRCA2 are the most important, but in Section A we list other genes that have been associated with BC risk. Current and future epidemiology is likely to change direction radically, and emphasise another class of genes, called polygenes. The onset of BC and OC is believed to be linked not only to the BRCA1/2 genes but also to a polygenic component. In fact, diseases associated with mutations in single genes alone are exceptional, the vast majority of genetic risks in adult life are almost certainly polygenic, and may be influenced by environment and lifestyle too (hence the name 'multifactorial disorder' which is often used to describe them). Even when major genes may cause a disease, it is possible that the majority of familial clustering of the disease may be caused by polygenes. This is very likely to be the case with BC (see Section 2.2.1). This epidemiological breakthrough will offer a completely new perspective on the insurance issues raised by knowledge of an individual's genetic profile.

Recently Antoniou *et al.* (2002) examined a number of genetic models for BC/OC risk, using data that included both high-risk families and families not selected for known BC risk. The best-fitting model incorporated BRCA1, BRCA2 and a polygene that modified the rates of onset of BC. Since their paper, and other published sources, allows all the $\mu_g(x)$ to be found, it is simple to build an actuarial model for CI insurance assuming genotypes to be known, hence to answer the question: what is the effect of pricing CI insurance if the polygene is allowed for, as well as the major genes BRCA1 and BRCA2? Put another way, how reliable is a genetic test that shows a BRCA1/2 mutation to be present, if the polygene is *not* taken into account? This bears directly on the criteria that GAIC has published for assessing the use of genetic tests by insurers.

2.2 The Model of Antoniou *et al.* (2002)

2.2.1 Breast Cancer and Polygenes

The risks of BC and OC onset were linked to mutations in the BRCA1 and BRCA2 genes in the 1990s, triggering a search for other genes implicated in tumour formation. This search is ongoing because it is estimated that BRCA1, BRCA2 and the other possible high-risk genes found to date (see Section A), account for only about 25% of the observed familial clustering (Struewing, 2004; Easton, 2005). Part of the problem is that mutations in BRCA1/2 are quite rare, their frequencies being estimated to be 0.051% and 0.068% respectively (Antoniou *et al.*, 2002).

It is widely believed that the remaining component arises from the combined influence of common alleles in several genes that each, individually, has only a small effect on the risk of BC. Such a configuration is called 'polygenic' (see Section 1.2.3) and the genes which contribute to it may collectively be called a 'polygene'. Although it is unlikely that a polygene explains all of the remaining 75% of the familial variation (there are other shared factors within families, such as diet and socio-economic status) it may explain a larger proportion than do any of the major genes.

Common low risk genes can account for significant proportions of disease in the population. Consider a disease with prevalence P_G in those with mutations in the gene G and prevalence P_0 in those without mutations. Now, given that gene G has a mutant allele that confers relative risk RR and has frequency q, the total prevalence of the disease is:

$$P = q\ P_G + (1-q)\ P_0 = q\ RR\ P_0 + (1-q)\ P_0. \tag{2.2}$$

Since the proportion of the disease in the mutation carrier population that is attributable to G is $(P_G - P_0)/P_G = (RR-1)/RR$, Levin (1953) proposes the following formula to estimate the proportion of the disease that may be attributed to the gene:

$$\frac{q\ RR\ P_0\ (RR-1)/RR}{q\ RR\ P_0 + (1-q)\ P_0} = \frac{q\ (RR-1)}{q\ (RR-1)+1}. \tag{2.3}$$

This is equivalent to the population attributable risk (PAR) statistic (or Levin's formula) which tells us the proportion of disease that would be wiped out if the exposure (in this case the gene) were eliminated. This equation has been recently advocated by Rebbeck (1999) for measuring the genetic contribution to cancer risk. If we consider a rare allele, say $q = 0.001$, with high risk, $RR = 20$, the proportion of disease attributable to this mutation (using Equation (2.3)) is just under 2%. However, if we consider a common ($q = 0.5$) low risk gene ($RR = 1.5$) the proportion attributable is 20%. The difference is quite large.

A more detailed account of the PAR is found in Hanley (2001). This work also includes a method for dealing with multiple types of exposure, which will be a useful tool for measuring the proportion of cancer that is attributable to polygenic risk.

2.2.2 The Hypergeometric Polygenic Model

The inheritance of major (single) genes, except those carried on the sex chromosomes, is usually assumed to follow Mendel's laws, as summarised in Section 1.1.6. Thus the chance that a child receives either copy of a gene carried by a given parent is 1/2. This is quite tractable if we are interested in a small number of major genes, each with a small number of alleles. For example if we regard BRCA1 and BRCA2 as each having two alleles (mutated and normal) there are only $3 \times 3 = 9$ possible genotypes, whose frequencies can be calculated exactly if the allele frequencies are known.

However a polygenic model may involve a large number of genes, each with several alleles. In principle Mendel's laws may still be applied, but the number of possible genotypes quickly becomes intractable in many practical problems. For example, if six genes contribute to the polygene, and each has two alleles, there are $3^6 = 729$ possible genotypes. It may be necessary (for example, in computing a likelihood) to form sums over all possible joint genotypes of all the family members.

Consequently, approximate models of the polygenic contribution to a disease have been proposed. A widely used assumption is that the polygenotype is represented by a numerical value on a continuous scale, and the distribution of these values in the population is Normal. This can be motivated by applying the Central Limit Theorem

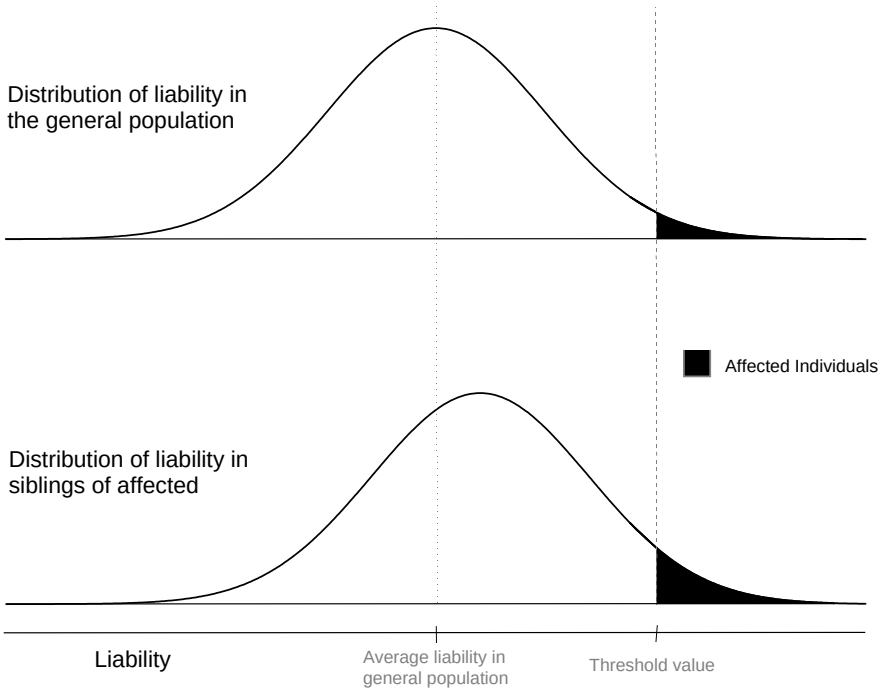


Figure 2.1: The polygenic threshold model of Falconer (1981). Individuals whose liability is above the threshold value are affected. On average, siblings of affected individuals have higher liability than the general population. Consequently more siblings exceed the threshold value for disease.

to a model in which total disease risk is the sum of the disease risks associated with all the alleles contributing to the polygenotype, with suitable independence assumptions. The polygenic disease risk may then be a suitable function of the polygene's numerical value, or the disease may be assumed to occur if the polygene's value exceeds a threshold. In respect of the latter, Falconer (1981) proposed a polygenic model for dichotomous conditions by postulating a variable and continuous underlying liability (possibly derived from a combination of genetic and environmental effects). This 'polygenic threshold model' (Figure 2.1) assumes that all individuals with liability above a certain threshold value will become affected by the disease.

While this gives a simple model of a polygene's effect, it makes it difficult to model inheritance. The question to be answered is: what is the conditional probability that the child of parents with known polygenotypes will have any given polygenotype? Passage to the continuous limit simplifies the problem of having many additive con-

tributions to the risk, but at the same time it turns the combinatorics of inheritance from hard to impossible.

As a result, when the inheritance of a polygene must be modelled, approximations may be made in the other direction, from continuous to discrete. The numerical polygenotype is assumed to take values in a discrete distribution with a suitable shape, for which there is a plausible model of transmission from parents to children. (Note that this discretisation does not mean a return to the Mendelian model; it is not now genes that are transmitted from parents to children but just a numerical 'value' representing the polygene.) Before giving an example, we fix terminology, by making the following conventions:

(a) The word 'polygene' will mean the collection of genes that constitute it — actual physical segments of DNA.

(b) Variants of a gene that contributes to a polygene will be called 'polygenic alleles'.

(c) The word 'polygenotype' means a numerical value representing the polygene.

Our example is the hypergeometric model of Lange (1997), derived from Cannings, Thompson & Skolnick (1978). It was used by Antoniou *et al.* (2002) to represent a polygenic component of BC risk, and it will be central to this chapter and the three that follow.

Suppose n genes, inherited independently of each other, contribute to the polygene, and each has an 'adverse' allele and a 'beneficial' allele which are equally common. An adverse allele contributes $+1/2$ to the numerical value of the polygenotype, and a beneficial allele $-1/2$. Since a person has two copies of each gene, the polygene is defined by the total number of adverse alleles, the possibilities being $0, 1, \ldots, 2n$. The corresponding numerical values of the polygenotype are $-n, -(n-1), \ldots, (n-1), n$, meant to suggest that 'negative' polygenotypes present below average risk, while 'positive' polygenotypes present above average risk. The mother's, father's and child's polygenotypes are random variables denoted P_m, P_f and P_c respectively. Assuming the parents to be sampled randomly from the population, their polygenotypes are independently binomially distributed with parameter $(2n, 1/2)$, for example:

$$\mathrm{P}[P_m = p_m] = \binom{2n}{p_m + n} \left(\frac{1}{2}\right)^{2n} \qquad \big(p_m = -n, -(n-1), \ldots, (n-1), n\big). \quad (2.4)$$

Thus the 'extreme' polygenotypes are uncommon, and the 'central' polygenotypes much more common. This is consistent with the Punnet square from Section 1.1 with n loci.

By the assumed independence the probability of the parents' joint polygenotype is:

$$\mathrm{P}[P_m = p_m, P_f = p_f] = \binom{2n}{p_m + n} \binom{2n}{p_f + n} \left(\frac{1}{2}\right)^{4n}. \quad (2.5)$$

Polygenes are transmitted from parents to children by independently sampling, without replacement, n polygenic alleles from the mother and n from the father. Conditional probabilities for an offspring's polygenotype are then:

$$\mathrm{P}[P_c = p_c | P_m = p_m, P_f = p_f] = \sum_{r=\max[0, p_c - p_f]}^{\min[p_m+n, p_c+n]} \frac{\binom{p_m+n}{r}\binom{n-p_m}{n-r}}{\binom{2n}{n}} \frac{\binom{p_f+n}{p_c+n-r}\binom{n-p_f}{r-p_c}}{\binom{2n}{n}}. \quad (2.6)$$

This is the convolution of two independent hypergeometric distributions representing the sum of the father's and the mother's contributions to their child's polygenotype. For further details see Lange (1997).

Furthermore, we can verify that independent samples of polygenotypes from the population are distributed binomially by noting that:

$$\sum_{p_m}\sum_{p_f} \mathrm{P}[P_c = p_c | P_m = p_m, P_f = p_f]\mathrm{P}[P_m = p_m, P_f = p_f] = \mathrm{P}[P_c = p_c], \qquad (2.7)$$

where P_c has the same distribution as P_m and P_f (see Equation (2.4)).

2.2.3 The Model of Antoniou *et al.* (2002)

Antoniou *et al.* (2002) fitted several alternative models to a set of high-risk families (each with multiple cases of BC or OC) and a set of unselected BC cases. The best-fitting model was a mixed major gene and polygenic model, in which the major genes were BRCA1 and BRCA2. The site of a mutation on BRCA1/2 was not considered; mutations were either present or absent. Previous studies have shown different mutation sites on the BRCA genes to display different risks of onset and aggressiveness after onset, but this aspect of the epidemiology of BC/OC is not yet developed enough to be taken into account.

For convenience, we use the term 'BRCA0 genotype' to indicate a person who carries neither BRCA1 nor BRCA2 mutations, and let 'BRCA1 genotype' and 'BRCA2 genotype' refer to mutation carriers, although strictly there is no such allele as BRCA0.

The authors used the national incidence rates for England and Wales in 1983–87 as baselines and estimated the relative risks of BC and OC in respect of BRCA1 and BRCA2 mutation carriers, piecewise constant over 10-year age groups between ages 30 and 69. These are shown in Table 2.1. Since they did not publish the baseline rates, we calculated our own using ONS statistics for England and Wales in 1983–87 and cancer registrations over the same period (ONS, 1999). These are shown in Figure 2.2, along with crude estimates of those used by Antoniou *et al.* (2002), obtained by dividing absolute onset rates by the relative risks. Thus we have onset rates $\mu^{BC}_{BRCAi}(x)$ and $\mu^{OC}_{BRCAi}(x)$, for $i = 0, 1, 2$.

Table 2.2 compares the BC incidence rates of BRCA1 and BRCA2 mutation carriers from this study with those of the earlier study by Ford *et al.* (1998) (the basis of

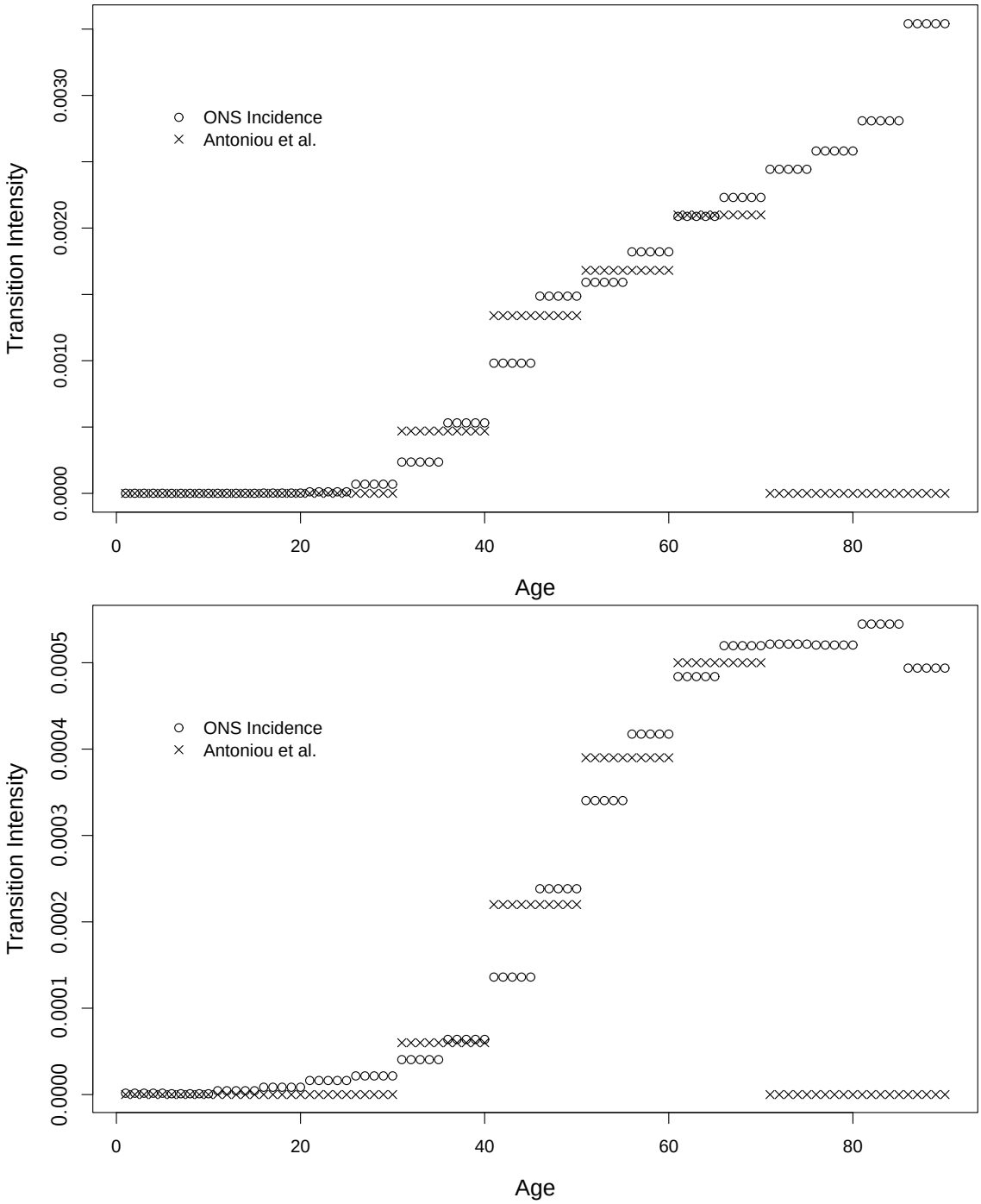


Figure 2.2: Baseline incidence rates for BC (top) and OC (bottom) from ONS figures for England and Wales (1983–1987) and figures from Antoniou *et al.* (2002)

Table 2.1: The relative risks for BC and OC BRCA1 or BRCA2 mutation carriers estimated by Antoniou *et al.* (2002). The baselines are the onset rates in England and Wales in 1983–87.

	Breast Cancer		Ovarian Cancer	
Age	BRCA1	BRCA2	BRCA1	BRCA2
30 – 39	23.88	17.52	3.43	3.67
40 – 49	12.40	10.80	53.32	2.00
50 – 59	4.91	12.11	20.86	11.85
60 – 69	2.31	12.53	19.51	8.32

Table 2.2: Comparison of the incidence rates for breast cancer estimated by Antoniou *et al.* (2002) and Ford *et al.* (1998).

	Antoniou *et al.*		Ford *et al.*	
Age	BRCA1	BRCA2	BRCA1	BRCA2
30 – 39	0.011222	0.008236	0.01618	0.0118
40 – 49	0.016621	0.014471	0.04749	0.0210
50 – 59	0.008255	0.020352	0.03480	0.0318
60 – 69	0.004843	0.026326	0.02162	0.1180

the actuarial model of Macdonald, Waters & Wekwete (2003a)). The trends with age are similar, but the rates from Antoniou *et al.* (2002) are much lower, particularly for older BRCA2 mutation carriers. This is as expected, because Ford *et al.* (1998) included only high-risk families (those with at least four cases of BC) whereas Antoniou *et al.* (2002) included a population-based cohort. Both studies focused on early onset of BC, with relatively few cases of onset at ages over 50–55, possibly leading to underestimated risk at higher ages.

The polygenotype is modelled as a Normal random variable R with mean 0 and standard deviation (the fitted parameter) $\sigma_R = 1.291$. It modifies the BC risk regardless of BRCA genotype as follows:

$$\mu^{BC}_{BRCAi}(x, R) = \mu^{BC}_{BRCAi}(x)\ e^R. \tag{2.8}$$

The Normal polygenic model was discretised to calculate likelihoods. Antoniou *et al.* (2002) used the hypergeometric model (see Section 2.2.2) with $n = 3$, thus seven polygenotypes P, with values $-3, -2, -1, 0, 1, 2, 3$, binomially distributed as in Equa-

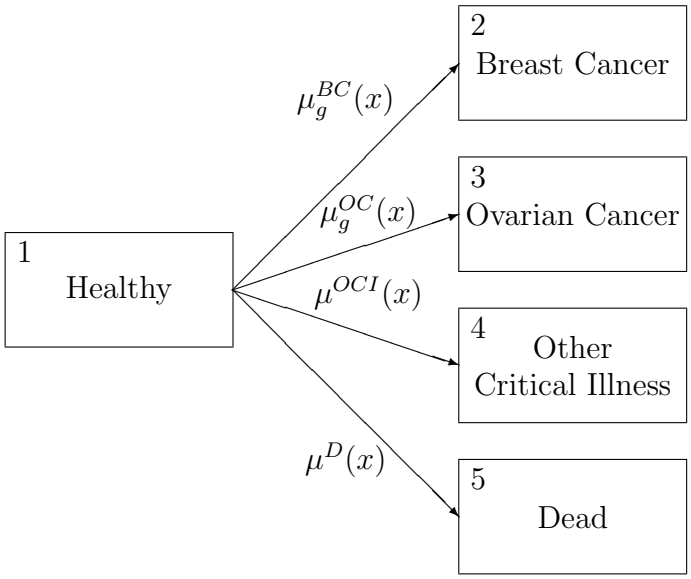


Figure 2.3: A model of the life history of a critical illness insurance policyholder, beginning in the Healthy state. Transition to the non-Healthy state d at age x is governed by an intensity $\mu^d(x)$ depending on age x or, in the case of BC and OC, $\mu_g^d(x)$ depending on genotype g as well.

tion (2.4). Values of R were approximated in terms of values of P (equating second moments) as follows:

$$R \approx \frac{P}{\sqrt{n/2}}\,\sigma_R. \tag{2.9}$$

The polygenotype did not affect the incidence of OC, or of any other disorder.

As we will need to model the transmission of polygenotypes from parents to children, we will use the same model.

2.3 A Model for Critical Illness Insurance

2.3.1 The Model

Figure 2.3 shows a continuous-time Markov model of a CI insurance contract. The transition intensities from 'Healthy' to 'Other Critical Illness' and 'Dead' are taken from Gutiérrez & Macdonald (2003). We provide some details of these intensities

in Section B. This is the model we will use to find the premiums payable for a stand-alone CI policy.

2.3.2 Premiums Based on Known Genotypes

Table 2.3 shows the net rates of level premium, payable continuously, for CI insurance cover at several entry ages and policy terms. The premium rates are expressed as a percentage of those for a woman who carries no BRCA1/2 mutation (genotype BRCA0) and who has the 'neutral' polygenotype $P = 0$, which we take to be the 'standard' premium. The force of interest is 0.05. Expected present values (EPVs) were found numerically by solving Thiele's equations (Hoem 1988) (see Section 1.4.1) using a Runge-Kutta algorithm with step size 0.0005 years.

Table 2.3: Level net premium for women, depending on polygenotype, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with the mean polygene $P = 0$.

Major Genotype	Polygenotype	Age 20				Age 30			Age 40		Age 50
		10 years	20 years	30 years	40 years	10 years	20 years	30 years	10 years	20 years	10 years
		%	%	%	%	%	%	%	%	%	%
BRCA0	−3	94.0	86.0	82.4	84.0	81.6	79.9	82.5	78.6	82.4	85.3
	−2	94.5	87.2	83.8	85.3	83.1	81.5	84.0	80.4	83.9	86.5
	−1	96.0	90.5	88.0	89.1	87.5	86.3	88.1	85.4	88.0	90.0
	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	+1	111.4	127.4	134.0	130.6	136.0	139.0	133.6	141.6	134.2	128.6
	+2	144.6	205.3	228.4	213.3	238.7	248.7	226.5	260.7	231.0	210.5
	+3	239.2	423.7	475.2	414.6	531.3	546.6	467.1	597.8	499.8	444.3
BRCA1	−3	94.0	102.4	182.4	167.7	106.7	201.2	178.7	263.6	205.1	154.9
	−2	94.5	126.2	203.6	182.4	143.0	227.4	196.0	285.3	218.5	160.9
	−1	96.0	193.9	263.4	223.9	246.9	302.0	245.6	347.2	256.8	177.9
	0	100.0	383.3	425.6	336.1	542.5	511.7	386.2	523.8	367.8	226.7
	+1	111.4	887.4	823.8	609.1	1372.0	1080.2	774.2	1020.8	691.5	366.1
	+2	144.6	2057.4	1575.5	1112.2	3605.9	2490.6	1760.2	2376.6	1642.8	762.6
	+3	239.2	3944.7	2464.5	1705.7	9068.7	5610.6	3958.7	5908.8	4271.7	1875.7
BRCA2	−3	94.0	99.4	95.2	107.5	102.2	95.4	109.1	91.2	111.1	129.2
	−2	94.5	116.9	113.2	123.4	128.8	117.1	127.2	110.3	127.4	143.9
	−1	96.0	166.9	164.1	167.6	205.1	179.0	178.3	164.9	173.8	185.8
	0	100.0	307.7	303.1	284.3	422.7	353.1	318.7	320.6	304.3	306.0
	+1	111.4	690.0	652.1	551.0	1036.9	824.7	677.7	760.0	658.7	647.9
	+2	144.6	1628.6	1347.1	1000.3	2718.3	1988.6	1480.4	1964.3	1553.9	1609.5
	+3	239.2	3373.8	2230.8	1547.9	6975.4	4482.6	3170.4	5101.4	3753.5	4293.2

In CI insurance, premiums in excess of 300% to 350% of the standard premium usually result in cover being declined. Many of the ratings for known BRCA1 and BRCA2 mutation carriers are above this level. Previous studies using quite recent epidemiology but the major genes only have reported that both BRCA1 and BRCA2 mutation carriers are likely to be declined for any combination of entry age and term (Gui *et al.*, 2006). Our results with the polygene $P = 0$ mostly agree with this.

The variation by polygenotype is the most striking feature of these results. And, since it affects the whole population, not just the carriers of rare mutations, it presents for the first time a widespread major variation of a genetic risk factor.

(a) The polygene alone (genotype BRCA0) leads to premiums for the highest risk ($P = +3$) that are up to 7.6 times those for the lowest risk ($P = -3$). Variation of this order caused by a major gene would probably be worthy of an actuarial study in its own right.

(b) In some instances a BRCA1/2 mutation carrier with a protective polygenotype may be eligible for a lower premium than non-mutation carriers with a risky polygenotype.

(c) We see that BRCA1/2 mutation carriers can be offered CI insurance at most entry ages and policy terms if they have a strongly protective polygenotype. Thus there is potential for genetic testing to make insurance more accessible under a lenient moratorium (one in which genetic test results may be disclosed if it is to the applicant's advantage).

On the other hand, premiums are even higher than previously reported for women with a detrimental combination of genotypes. The premium rate in the worst case (polygenotype +3 and major genotype BRCA1) is over 90 times the 'standard' rate and up to 85 times the premium rate for a BRCA1 mutation carrier with polygenotype -3. For BRCA2 mutation carriers the corresponding multiples are about 70 and 68 times.

2.3.3 An Australian Population

The Antoniou *et al.* (2002) study used two cohorts of BC pedigrees, one group of 1,484 families from the Anglican Breast Cancer (ABC) study and one group of 156 families with multiple BC cases. Another study by Cui *et al.* (2001) was conducted alongside the Antoniou *et al.* (2002) study using Australian pedigrees.

Cui *et al.* (2001) studied families ascertained through 858 women diagnosed with BC before the age of 40. They fitted the same polygenic model of BC risk as Antoniou *et al.* (2002). In the Australian study σ_R was estimated as 1.533. Although Cui *et al.* (2001) fitted a major gene effect with the polygene, they only included one possible susceptibility locus. Since this lacked resemblance to the BRCA1 and BRCA2 genes we excluded it and concentrated on the fit which included only the polygene. As a result we consider the fit which uses only the 824 families with no known BRCA1 or BRCA2 mutations.

We display the intensities of BC and OC in Australia in Figure 2.4 for comparison with that of the UK (Figure 2.2). The incidences of BC for the Australian population were taken from the same source as Cui *et al.* (2001): the Australian Institute for Health and Welfare (1999). The breast cancer incidence for the period 1982–1996 was divided between three periods: 1982–1986, 1987–1991 and 1992–1996. We averaged the incidence over each of these periods for each age and then used kernel-smoothing (with a bandwidth of 10 years) to obtain an estimate of the incidence rate in the period 1982–1996. We obtained the population incidence rate of ovarian cancer from 1998 estimates by the National Breast Cancer Centre (2002).

The premium rates for each polygenotype are given in Table 2.4. Note that these calculations were made using the the CI model (Figure 2.3) with intensities of morbidity and mortality based on UK experience. Hence, Table 2.4 is given only as a basic guide to the premiums required under a different parameterisation of the polygenic model.

Because of the higher estimate of σ_R more dispersion of risk is conferred by the polygene and hence the premiums in Table 2.4 show greater variation than those in Table 2.3 (for BRCA0 individuals). However since we consider the polygene to be the

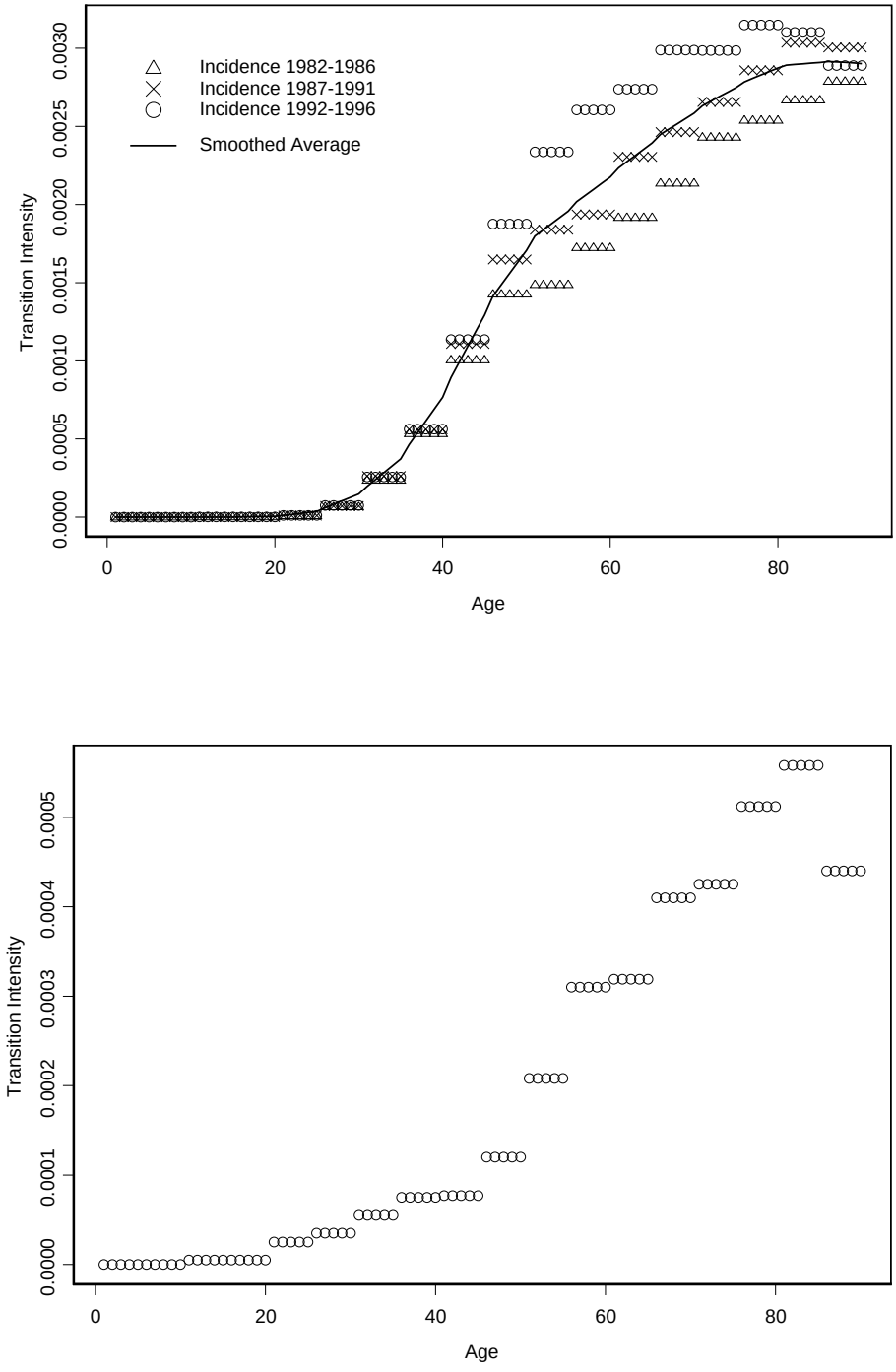


Figure 2.4: Baseline incidence rates for BC (top) and OC (bottom) from the Australian Institute of Health and Welfare (1999) and the National Breast Cancer Centre (2002), respectively

only genetic influence on BC onset in the Australian population, the estimate of σ_R most likely compensates for the lack of major gene effect and is greater than if major genes were considered in the model.

Table 2.4: Level net premium for women free of BRCA1/2 mutations, depending on polygenotype, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with the mean polygene $P = 0$. Based on an Australian population.

	Age 20				Age 30			Age 40		Age 50
Polygenotype	10 years	20 years	30 years	40 years	10 years	20 years	30 years	10 years	20 years	10 years
	%	%	%	%	%	%	%	%	%	%
−3	94.0	85.7	82.0	83.6	81.3	79.5	82.2	78.2	82.1	85.0
−2	94.3	86.5	83.1	84.6	82.4	80.7	83.3	79.5	83.1	85.9
−1	95.5	89.6	86.9	88.1	86.3	85.0	87.0	84.1	86.9	89.1
0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
+1	115.2	136.6	145.3	140.6	148.0	151.9	144.7	155.6	145.6	138.2
+2	169.0	262.7	295.7	270.3	314.9	328.3	292.4	348.7	302.0	271.3
+3	356.4	684.3	738.2	607.8	890.0	884.9	721.6	1007.6	816.5	731.9

2.3.4 A Comment on Genetic Tests for Polygenotypes

References to 'known' polygenotypes should not lead readers to suppose they might soon be detected by DNA-based genetic tests. Our model of a polygenotype is a numerical value, whereas a real polygenotype is a combination of (possibly very many) alleles. In order to test for a polygene and relate the result to a risk estimate, all the complications that drive geneticists to use the simplified model will have to be overcome. Moreover, it seems unlikely that genetic risks will be capable of being understood in isolation, but only in combination with other major risk factors.

2.4 Comparison of Data and Methods

Gui *et al.* (2006) also studied the genetics of BC and OC to determine how the insurance industry may be affected by genetic testing. Since their work is of a similar nature to ours, in this section we will compare their parameterisations with that of ours, so that we may gauge how much it is possible to compare results.

2.4.1 The Baseline Hazard

The baseline cancer incidence rates used in the Gui *et al.* (2006) study are calculated from the cancer registry of England and Wales between 1973 and 1977. Although we use the same cancer registry, the period of observation used in our figures is between 1983 and 1987. This inconsistency was unavoidable given that these intervals were selections made by the study groups that the work was based on. Gui *et al.* gathered their intensities from a model of Antoniou *et al.* (2003) and, as previously stated, we worked from the model of Antoniou *et al.* (2002).

Between 1971 and 2003, the age-standardised incidence of cancer in females increased by around 40% (National Statistics Online). As BC and OC are respectively the first and fourth most common female cancers, it is very likely that the incidence of these two cancers has risen. In fact, by determining the incidences in each of the two periods (1973–1977 and 1983–1987) from the cancer statistics (ONS, 1999) we can see in Figure 2.5 that there has been a substantial increase in incidence in the

later period (1983–1987). It is likely that at least some of this increase in incidence can be attributed to developments in screening and diagnostic technologies. This increase (whether superficial or not) means that the baseline data used in our work is reasonably higher than that used in Gui *et al.* (2006). All else equal, this would imply that our calculation of a standard CI policy premium, without any available genetic information, would be higher than that of Gui *et al.*.

2.4.2 Relative Risks For BRCA1/2 Mutation Carriers

The relative risk of a BRCA1/2 mutation acts multiplicatively on the baseline transition intensity. The relative risks (in 10-year age groups) were estimated for ages 30–69 by Antoniou *et al.* (2002) and for ages 20–69 by Antoniou *et al.* (2003).

The different relative risks used in the two studies are displayed in Table 2.5. We can see that the Gui *et al.* relative risks (those of Antoniou *et al.* (2003)) are somewhat greater than those of Antoniou *et al.* (2002) for nearly all categories apart from the relative risk of BC in BRCA2 mutation carriers.

Table 2.5: The relative risks of BC and OC for BRCA1/2 mutation carriers determined by Antoniou *et al.* (2002) and by Antoniou *et al.* (2003) in 10-year age intervals.

	Antoniou *et al.* (2002)				Antoniou *et al.* (2003)			
Age	Breast Cancer		Ovarian Cancer		Breast Cancer		Ovarian Cancer	
	BRCA1	BRCA2	BRCA1	BRCA2	BRCA1	BRCA2	BRCA1	BRCA2
20 – 29	N/A	N/A	N/A	N/A	17.0	19.0	1.0	1.0
30 – 39	23.88	17.52	3.43	3.67	33.0	16.0	49.0	1.0
40 – 49	12.40	10.80	53.32	2.00	32.0	9.9	68.0	6.3
50 – 59	4.91	12.11	20.86	11.85	18.0	12.0	31.0	19.0
60 – 69	2.31	12.53	19.51	8.32	14.0	11.0	50.0	8.4

2.4.3 Penetrance

The penetrance of a gene is an individual's cumulative probability of developing the disease associated with the gene at some given time, conditional on the fact that the individual carries the mutation in that gene. For example, the penetrance of

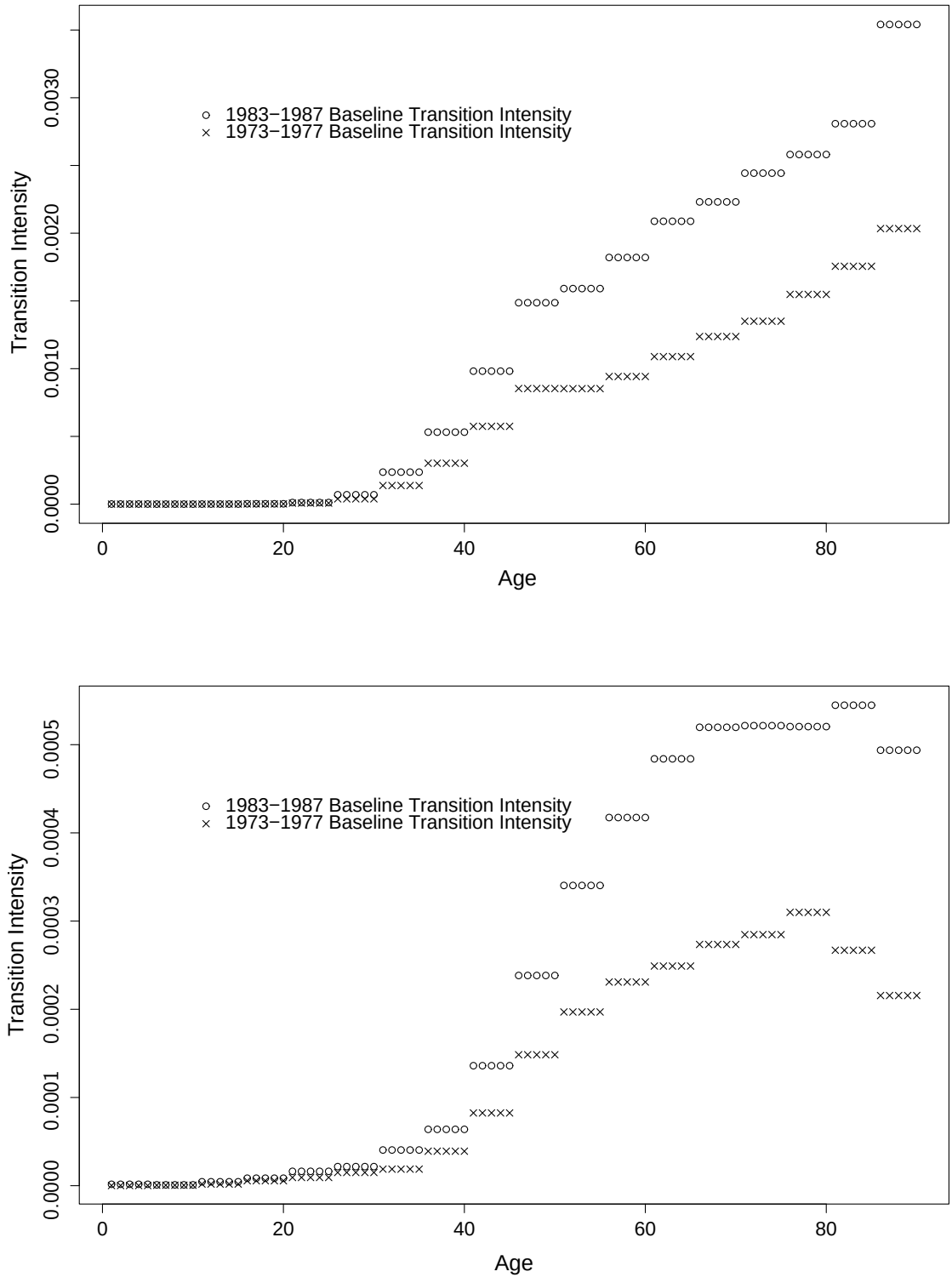


Figure 2.5: Baseline incidence rates for BC (top) and OC (bottom) from ONS figures for England and Wales (1983–1987) and (1973–1977)

the BRCA1 gene is derived from the incidence of BC given BRCA1 mutation status, $\mu^{BC}_{BRCA1}(x)$, by:

$$q^{BRCA1}(x) = 1 - \exp\left(-\int_0^x \mu^{BC}_{BRCA1}(s)ds\right). \tag{2.10}$$

This is equivalent to the right-hand side of Equation (1.5), the cumulative probability of death or onset in a single-decrement model.

The penetrances at ages 50 and 70 are given in Table 2.6. We see much higher penetrance estimates in the 2003 study except for BRCA2 mutation carriers for BC (as we saw with the relative risks).

Table 2.6: The penetrances, $q^g(x)$, for BC and OC by age 50 and 70 for BRCA1/2 mutation carriers determined by Antoniou *et al.* (2002) and Antoniou *et al.* (2003).

Age	Antoniou *et al.*	Breast Cancer		Ovarian Cancer	
		BRCA1	BRCA2	BRCA1	BRCA2
		%	%	%	%
50	(2002)	24.3	20.3	11.3	0.7
50	(2003)	38.3	16.2	13.2	1.2
70	(2002)	35.3	50.3	25.9	9.1
70	(2003)	65.0	44.7	39.1	11.1

In Table 2.7 we give the level premiums for a CI policy that we calculated for BRCA1 and BRCA2 mutation carriers (which is polygenotype category 0 in Table 2.3) alongside the corresponding premiums calculated by Gui *et al.* (2006). The differences in penetrances explain the differing results we see for premium rates between our work and theirs. For all entry ages and terms the Gui *et al.* BRCA1 mutation carrier premiums are significantly higher than ours. For BRCA2 mutation carriers however, the premiums are very similar throughout.

Table 2.7: Level net premiums for CI cover as a percentage of standard risks, for BRCA1 and BRCA2 mutation carriers. Figures in brackets are the premiums from Gui *et al.* (2006) using 100% incidence rates.

Major	Age 20				Age 30			Age 40		Age 50
Genotype	10 years	20 years	30 years	40 years	10 years	20 years	30 years	10 years	20 years	10 years
	%	%	%	%	%	%	%	%	%	%
BRCA1	100.0	383.3	425.6	336.1	542.5	511.7	386.2	523.8	367.8	226.7
	(977)	(1176)	(905)	(682)	(1347)	(967)	(725)	(842)	(654)	(532)
BRCA2	100.0	307.7	303.1	284.3	422.7	353.1	318.7	320.6	304.3	306.0
	(366)	(416)	(361)	(317)	(449)	(369)	(323)	(338)	(308)	(296)

2.4.4 Mutation Frequencies

The genotype mutation frequencies for BRCA1 and BRCA2 determined by Antoniou *et al.* (2002) are based on population frequency estimates derived from an analysis with a polygenic component. The mutant allele frequencies for BRCA1 and BRCA2 respectively are:

$$p_1 = 0.00051,$$

$$p_2 = 0.00068.$$

The mutant allele frequencies used in Gui *et al.* were found without accounting for a polygenic component. Because the polygenic component explains a portion of the familial risk, the allele frequencies were larger, on aggregate, in the study which did not include it, *i.e.* BRCA1/2 mutations are estimated to be less widespread in Antoniou *et al.* (2002) since the polygene is doing some of the work of making BC hereditary. Under the sporadic model in Antoniou *et al.* (2003) (a model with no extra genetic component besides BRCA1/2) the mutant allele frequencies are:

$$p_1 = 0.000583,$$

$$p_2 = 0.000676,$$

which were those used in Gui *et al.*.

The allele frequencies do not actually tell us the proportion of the population who carry the mutated genes. To find these proportions we must calculate the *carrier* frequencies. Since each individual carries two independent copies of each gene, to be called a mutation carrier they either possess two of the mutated alleles (which is very unlikely in the case of BRCA1/2) or one mutated and the other non-mutated, which may happen in one of two ways. Thus, the carrier frequency, found from the allele frequency, p, is:

$$p^2 + 2p(1-p) \approx 2p(1-p). \tag{2.11}$$

The right-hand side of Equation (2.11) is perhaps the most realistic for the case of the BRCA1/2 alleles, since it is believed that carriers of homozygous mutations in either gene are non-viable.

These carrier frequencies become important in the next section, when we calculate premiums for individuals with a family history of BC or OC.

Chapter 3

Modelling Family History with the Polygenic Model

3.1 Introduction

3.1.1 Modelling Family History

We wish to study how the polygene affects CI insurance pricing if, as usual, only the existence of a family history is known. Previous studies have used Equation (2.1) directly, because a small number of major genes defines a small number of genotypes g. This is not the case with the polygenic model, in particular the conditional genotype probabilities in Equation (2.1) are intractable. We therefore simulate a large number of nuclear families, and assume that the children of these families make up the pool of potential applicants for insurance. The empirical distribution of genotypes in this simulated sample provides the probabilities in Equation (2.1) directly.

The problem is to find EPVs given a family history at age x , as in Equation (2.1). Assuming the genotype-specific onset rates to be known, this reduces to estimating the conditional probabilities:

$$\text{P[Genotype is } g \mid \text{Family history exists at age } x \text{]}. \tag{3.1}$$

First, we must define what is meant by 'family history'. That done, the calculation must be anchored by the assumption that some ancestors of the applicant have genotypes that are randomly and independently sampled from the distribution of genotypes in the population. We will assume this to be true of the applicant's parents; thus their genotype probabilities are known. Together with the transmission probabilities that govern the inheritance of genes, this fixes the genotype probabilities of the applicant and all her siblings. For every possible joint genotype of the entire family, we know the probabilities of critical illnesses, including BC and OC, striking before any given age, hence the probability of a family history arising. At this point, the computation of the probability (3.1) has become, in principle, just an application of Bayes' Theorem. However the summation is not over the applicant's possible genotypes as in Equation (2.1), but over all possible joint genotypes of the whole family.

The procedure outlined above was followed by Macdonald, Waters & Wekwete (2003a) for several definitions of family history. They also considered the more realistic possibility that the insurer may not have any information about the unaffected relatives of the applicant. Their approach could not be extended to a model of the insurance market, necessary to study the potential costs of adverse selection, because it did not model the development of a family history over time as a factor that might influence the decision to buy insurance or to take a genetic test. That step was taken by Gui *et al.* (2006) who pointed out that if the definition of 'family history' is such that at any given time it is either certainly present or certainly absent, the time at which it appears can be modelled as an event time in the usual framework of survival models, and the procedure outlined above can be modified to give an age-dependent 'rate of onset' of a family history. But this approach still depended on applying Mendelian transmission probabilities to just two major genes.

The polygenotype introduces a non-Mendelian model of transmission, which is not a real problem, and greatly increases the number of genotypes, which is. Thus we have chosen to estimate the probabilities (3.1) by simulation.

3.1.2 Definition of Family History

Wekwete (2002) set out the underwriting guidelines for applicants with a history of BC which at the time were used by three different UK insurance companies. They are reproduced in Table 3.1. Two observations can be drawn from these examples:

(a) Applicants over age 50 are not 'rated up' at all by Company A. Also, the rating up on applicants who have relatives affected before age 50 is much more severe than the rating up of applicants who have relatives above age 50.

(b) Declinature of applicants only occurs when the applicant has two or more affected relatives.

These observations help us to formulate a definition of a family history.

Our definition of a family history is based on a typical underwriting threshold, namely two first-degree relatives (FDRs, meaning parents and siblings) suffering onset of BC or OC before age 50. Under many underwriting standards this condition would lead to an extra premium being charged (Macdonald, Waters & Wekwete, 2003b). Note that this is quite different from clinical practice, in which a family history may be defined by a much more complex pedigree including second-degree and other relatives. To a clinician, also, a family history is defined by the circumstances of each patient. Thus we rely on the much simpler notion used by insurers.

3.2 Simulating Families

3.2.1 The Simulation Model

The approach is as follows:

(a) A family starts with two parents, whose major genotypes and polygenotypes are independently sampled from their respective distributions in the population, except that we disregard the probability that either parent has more than one mutation. This is consistent with the treatment of BRCA1 and BRCA2 in Antoniou *et al.* (2001). It is widely assumed by epidemiologists that a foetus with

Table 3.1: An example of CI underwriting procedure for BC family histories. Source: Wekwete (2002)

Age of Applicant	Number of Affected Relatives	Age at Diagnosis or Death	Rating Offered by Company A	Company B	Company C
≤ 40	1	< 50	+150	+100	+0
		50–64	+50	+0	+0
		> 65	+0	+0	+0
	2	< 50	Decline	Decline	+50
		50–64	+150	+50	+0
		> 65	+150	+0	+0
	>2	< 50	Decline	Decline	CMO
		50–64	+150	+50	+50
		> 65	+150	+0	+50
41–50	1	< 50	+100	+100	+0
		50–64	+0	+0	+0
		> 65	+0	+0	+0
	2	< 50	Decline	Decline	+50
		50–64	+100	+50	+0
		> 65	+100	+0	+0
	>2	< 50	Decline	Decline	CMO
		50–64	+100	+50	+50
		> 65	+100	+0	+50
> 50	1	< 50	+0	+100	+0
		50–64	+0	+0	+0
		> 65	+0	+0	+0
	2	< 50	+0	Decline	+50
		50–64	+0	+50	+0
		> 65	+0	+0	+0
	>2	< 50	+0	Decline	CMO
		50–64	+0	+50	+50
		> 65	+0	+0	+50

CMO: Refer to Chief Medical Officer.

Table 3.2: Distribution of the number of daughters born in a family. Source: Macdonald, Waters & Wekwete (2003a)

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

two mutations of the same BRCA gene will not be viable and will miscarry. We use the BRCA1 and BRCA2 mutation frequencies from the polygenic model in Antoniou *et al.* (2002), 0.051% and 0.068% respectively.

(b) The number of daughters the parents have is randomly sampled from a suitable distribution. We use that of Macdonald, Waters & Wekwete (2003a) which is given in Table 3.2. Hence the family size may vary from three to nine members, and the father is the only male. For simplicity, we assume that the mother has her children when she is age 30 and all daughters are the same age.

(c) Each daughter, independently of the others, inherits the major genes at random according to Mendel's laws, and the polygenotype at random according to Equation (2.6). We discard any family in which a daughter inherits any two major gene mutations, for the reasons given in (a).

(d) The life histories of the mother and daughters, in respect of the model in Figure 2.3, are simulated using a competing risks approach. We ignore male BC, and we assume that the mother is healthy at age 30.

After simulating a large number of such families, we can observe, at every age $x > 0$, the distribution of the genotypes of daughters in families in which a family history has appeared. We will describe this in Section 3.2.4.

3.2.2 Simulating Competing Risks

There are four decrements in the model in Figure 2.3. Define T^{id} to be the random time at which the ith person in the simulated sample suffers decrement d, as if it

acted alone. In the simulation, the ith person's genotype is known, say it is g. Then T^{id} has distribution function, denoted $F_g^d(t)$, given by:

$$F_g^d(t) = 1 - \exp\left(-\int_0^t \mu_g^d(x+s)ds\right) \tag{3.2}$$

for $t < \infty$, possibly with a probability mass at $t = \infty$. This, and its inverse, can be computed and tabulated. The random variable $F_g^d(T^{id})$ is uniformly distributed on $[0, 1]$ so we simulate a uniform $[0, 1]$ random variable, denoted a^{id}, and solve numerically the equation $F_g^d(t^{id}) = a^{id}$, to obtain our simulated value t^{id}. The ith person's life history is then represented by the pair (t^i, d^i) where $t^i = \min[t^{i1}, t^{i2}, t^{i3}, t^{i4}]$ and d^i is that decrement for which $t^{ij} = t^i$.

Note that each decrement in the model censors the others, so it is not possible for a woman who survives a heart attack (for example) to develop BC/OC subsequently. The effect is minimal at those ages where onset would contribute to a family history; by age 50 only about 6% of women have developed one of the other CIs.

3.2.3 Sampling Insurance Applicants from Simulated Families

We simulated 10,000,000 families as described above, containing in total 16,019,834 daughters. At any age x, those daughters still healthy constitute the pool of potential applicants for insurance. We assume that the insurer, in effect, samples randomly from this pool, knowing only whether each applicant has a family history or not. As well as using the maximum possible amount of information in the simulated families, this sampling scheme accounts correctly for the fact that there are more potential applicants than there are family histories; in larger families the appearance of a family history will affect more than one healthy daughter.

3.2.4 Applicant's Genotype Distribution

We can now estimate by direct enumeration the distribution of the applicant's genotype conditional on the observed family history. All applicants are healthy, but some

have a family history and others do not. This is all the insurer knows. We, however, also know into which of the following categories each applicant falls.

(a) Applicant is in a BRCA0 family (no mutations) and has BRCA0 genotype.

(b) Applicant is in a BRCA1 family and has BRCA0 genotype.

(c) Applicant is in a BRCA1 family and has BRCA1 genotype.

(d) Applicant is in a BRCA2 family and has BRCA0 genotype.

(e) Applicant is in a BRCA2 family and has BRCA2 genotype.

Table 3.3 shows the numbers of daughters who have no family history at selected ages from 0 to 60 years, grouped into the five categories above and the state occupied in the CI model (Figure 2.3). Table 3.4 shows the corresponding distribution of daughters who do have a family history. In both tables the potential insurance applicants are those in the Healthy state.

We further subdivide the numbers in Tables 3.3 and 3.4 by polygenotype. The results are too extensive to tabulate, so for illustration Figures 3.1 and 3.2 show histograms of the polygenotype distribution among healthy daughters with a family history, for the five major gene categories above, and ages 30 and 40 (Figure 3.1) and 50 and 60 (Figure 3.2). Note that few mutation carriers have a family history by age 30. This is because mutation carriers are rare, and before age 30 they share the population onset rates of BC and OC.

For brevity here we omit the polygenotype distributions of daughters with no family history (see Section 3.2.6). They are slightly more inclined to less risky values because carriers of more dangerous polygenotypes are more likely to have FDRs with risky polygenotypes, hence have a higher risk of developing a family history. This is most pronounced in BRCA2 mutation carriers, because the deleterious effects of BRCA2 mutations are relatively late-acting.

These empirical distributions (at all ages x, not just the selected ages illustrated) provide the conditional probabilities we need (Equation (3.1)) to calculate premiums for a daughter with a family history.

Table 3.3: Numbers of daughters with no family history and given major genotype, in each state in the CI model (see Figure 2.3), at selected ages.

Genotype			Daughters' Ages						
Family	Applicant	State	0	10	20	30	40	50	60
BRCA0	BRCA0	Healthy	15,944,331	15,888,102	15,828,825	15,686,533	15,265,549	14,140,830	12,290,717
BRCA1	BRCA0	Healthy	16,091	16,018	15,961	15,815	15,150	13,463	11,677
BRCA1	BRCA1	Healthy	16,045	15,859	15,799	15,655	12,815	8,857	6,817
BRCA2	BRCA0	Healthy	21,808	21,697	21,599	21,422	20,642	18,596	16,188
BRCA2	BRCA2	Healthy	21,559	21,380	21,273	21,046	18,014	13,879	9,850
BRCA0	BRCA0	BC	0	1,725	2,241	16,279	139,815	462,864	900,578
BRCA1	BRCA0	BC	0	25	25	43	122	323	742
BRCA1	BRCA1	BC	0	112	113	133	1,837	2,708	3,350
BRCA2	BRCA0	BC	0	33	34	51	161	440	989
BRCA2	BRCA2	BC	0	118	118	136	2,055	3,544	5,934
BRCA0	BRCA0	OC	0	262	1,287	4,195	11,900	37,058	86,850
BRCA1	BRCA0	OC	0	1	3	9	15	37	98
BRCA1	BRCA1	OC	0	23	24	25	34	779	1,396
BRCA2	BRCA0	OC	0	2	3	7	14	41	95
BRCA2	BRCA2	OC	0	3	7	12	44	72	586
BRCA0	BRCA0	Other	0	10,947	41,283	125,342	351,838	929,402	2,094,960
BRCA1	BRCA0	Other	0	17	52	136	367	926	2,065
BRCA1	BRCA1	Other	0	10	48	136	337	732	1,387
BRCA2	BRCA0	Other	0	11	57	162	496	1,238	2,800
BRCA2	BRCA2	Other	0	11	67	208	511	1,137	2,094
BRCA0	BRCA0	Dead	0	43,295	70,695	111,829	162,694	252,925	449,974
BRCA1	BRCA0	Dead	0	30	50	87	133	224	391
BRCA1	BRCA1	Dead	0	41	61	89	162	220	346
BRCA2	BRCA0	Dead	0	65	115	162	237	358	601
BRCA2	BRCA2	Dead	0	47	94	153	210	300	468

Table 3.4: Numbers of daughters with a family history and given major genotype, in each state in the CI model (see Figure 2.3), at selected ages.

Genotype			Daughters' Ages						
Family	Applicant	State	0	10	20	30	40	50	60
BRCA0	BRCA0	Healthy	0	0	0	80	5,936	48,761	35,837
BRCA1	BRCA0	Healthy	0	0	0	1	234	799	666
BRCA1	BRCA1	Healthy	0	0	0	6	148	427	288
BRCA2	BRCA0	Healthy	0	0	0	4	175	763	633
BRCA2	BRCA2	Healthy	0	0	0	3	134	497	244
BRCA0	BRCA0	BC	0	0	0	62	6,180	65,908	74,279
BRCA1	BRCA0	BC	0	0	0	0	58	233	284
BRCA1	BRCA1	BC	0	0	0	1	687	1,938	2,004
BRCA2	BRCA0	BC	0	0	0	0	70	287	357
BRCA2	BRCA2	BC	0	0	0	1	573	2,049	2,237
BRCA0	BRCA0	OC	0	0	0	11	278	2,869	3,342
BRCA1	BRCA0	OC	0	0	0	0	4	17	21
BRCA1	BRCA1	OC	0	0	0	0	13	322	364
BRCA2	BRCA0	OC	0	0	0	0	3	20	22
BRCA2	BRCA2	OC	0	0	0	0	9	37	63
BRCA0	BRCA0	Other	0	0	0	0	99	2,947	6,411
BRCA1	BRCA0	Other	0	0	0	0	5	55	124
BRCA1	BRCA1	Other	0	0	0	0	8	47	70
BRCA2	BRCA0	Other	0	0	0	0	5	49	101
BRCA2	BRCA2	Other	0	0	0	0	5	28	57
BRCA0	BRCA0	Dead	0	0	0	0	42	767	1,383
BRCA1	BRCA0	Dead	0	0	0	0	3	14	23
BRCA1	BRCA1	Dead	0	0	0	0	4	15	23
BRCA2	BRCA0	Dead	0	0	0	0	5	16	22
BRCA2	BRCA2	Dead	0	0	0	0	4	16	26

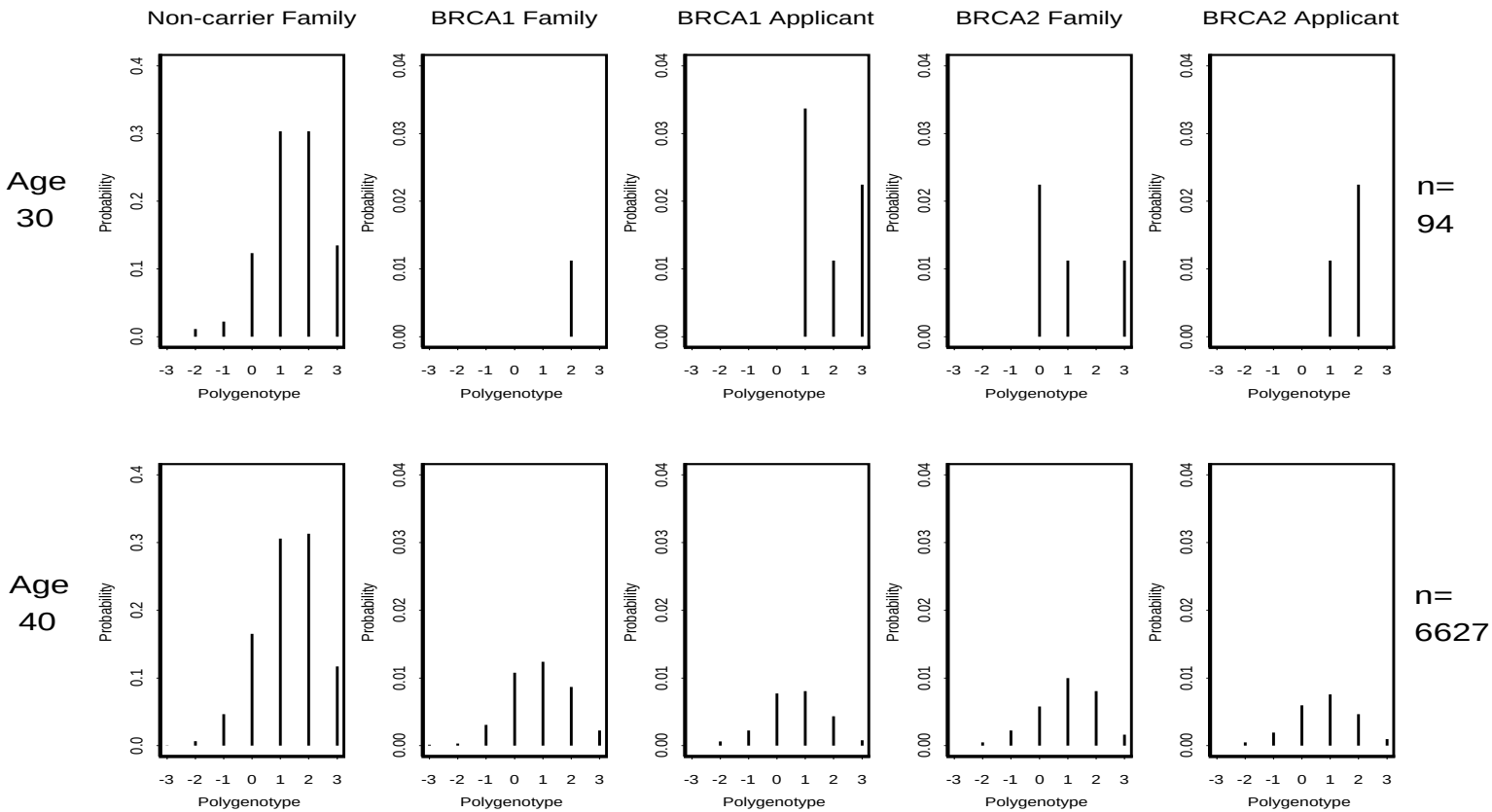


Figure 3.1: The distribution of polygenotypes by major genotype among healthy daughters aged 30 and 40, with a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families.

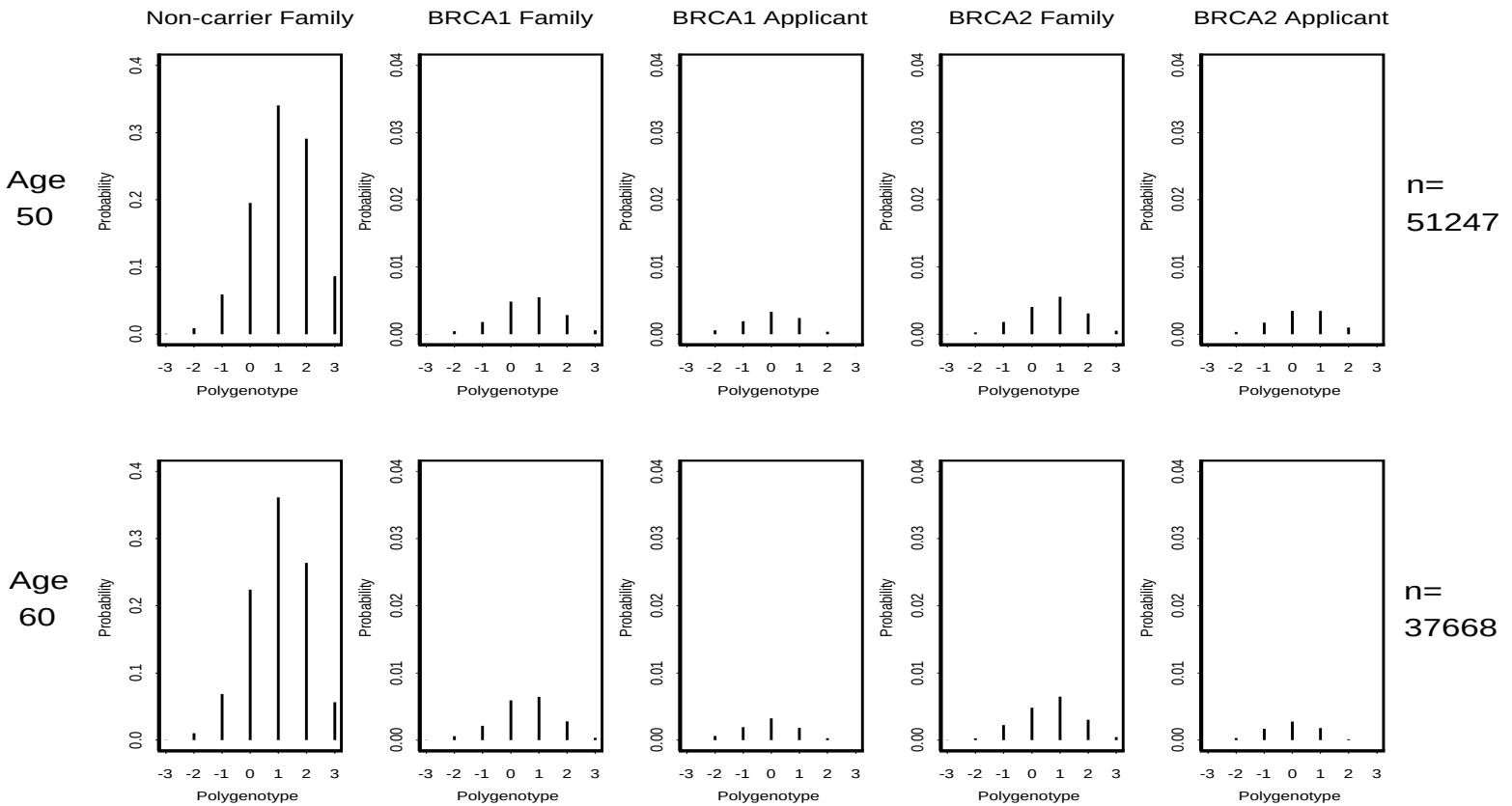


Figure 3.2: The distribution of polygenotypes by major genotype among healthy daughters aged 50 and 60, with a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families.

Table 3.5: Level net premium for females with a family history of BC or OC, as a percentage of the level net premium for a woman free of BRCA1/2 mutations and with polygenotype $P = 0$. The P + MG model uses both major gene and polygene probabilities in the weighted average EPVs, while the MG model uses only the major gene probabilities.

Definition of Family History	Genetic Model	Age 30			Age 40		Age 50
		10 yrs %	20 yrs %	30 yrs %	10 yrs %	20 yrs %	10 yrs %
1 Affected FDR	P + MG	183.2	176.4	160.6	166.2	151.1	138.2
	MG	105.4	104.5	103.3	102.8	102.0	101.0
2 Affected FDRs	P + MG	444.0	341.0	274.7	244.2	207.4	170.6
	MG	137.5	132.0	122.7	112.9	108.9	102.8
3 Affected FDRs	P + MG	100.0	100.0	100.0	410.7	314.1	215.6
	MG	100.0	100.0	100.0	148.2	134.3	106.8
4 Affected FDRs	P + MG	100.0	100.0	100.0	934.3	637.1	260.8
	MG	100.0	100.0	100.0	207.9	173.2	112.1

3.2.5 Premiums for an Applicant with a Family History

Sample level premiums for a daughter with a family history, applying for level CI insurance, are shown in the third line of Table 3.5. They are expressed as percentages of the relevant premium for a woman with major gene BRCA0 and polygenotype $P = 0$.

Table 3.5 shows the effect of allowing for or ignoring the polygene. The full model, labelled 'P + MG', uses both polygene and major gene probabilities in weighting EPVs. The major-gene-only model, labelled 'MG', uses only the major gene probabilities, assuming everyone has polygenotype $P = 0$. The latter are very much lower, but this has to be interpreted with care.

(a) The major-gene-only model is *not* comparable with previous actuarial studies of CI insurance that were based on the major genes only. Here it just isolates the contribution of the major genes to the familial risk, in the full model. The earlier studies were based on genetic models in which 100% of the familial risk was attributed to the major genes.

(b) What these figures *do* show, on comparison with the earlier studies, is that the larger proportion of the genetic risk of BC/OC lies with the polygene, not with

Table 3.6: Level net premium for females with a family history of BC or OC, as a percentage of the standard premium. The polygenic model is compared with the major-gene-only model of Gui *et al.* (2006). The latter assumed that onset rates of BC and OC among BRCA1/2 mutation carriers were either 100% or 50% of those estimated, as a rough allowance for ascertainment bias.

Definition of	Genetic		Age 30			Age 40		Age 50
Family History	Model	10 yrs	20 yrs	30 yrs	10 yrs	20 yrs	10 yrs
		%	%	%	%	%	%
2 Affected FDRs	P + MG	444	341	275	244	207	171
	MG	138	132	123	113	109	103
Gui *et al.* (2006)	100%	330	251	204	208	174	142
	50%	217	179	156	154	139	120

the major genes. This is a very significant conclusion, because genetic testing for the major genotypes is common, but there is no immediate prospect of defining and testing for polygenotype.

Under the major-gene-only model, policies taken out at age 20 have almost no additional risk because the probability of having developed a family history by age 20 is almost zero (which is consistent with Figure 3.1).

The premium increases shown under the full polygenic model (P + MG) range from 70–345%. The insurer probably would charge an extra premium given these results, and in some cases decline applicants. Clearly, this is a consequence of the definition of family history. We would expect stricter definitions to pinpoint the presence of major genes more accurately, though in a much reduced number of families. Table 3.5 also shows the increased premiums if a family history is defined as at least three or as at least four first-degree relatives with BC or OC before age 50. As expected, they are much higher, in some cases approaching the limit of insurability. However, such family histories are so rare before age 30, even among 10,000,000 simulated families, that the additional premiums were zero for policies taken out at that age.

Macdonald *et al.* (2003b) and Gui *et al.* (2006) gave premium ratings for CI insurance in the presence of a family history of BC or OC. Both used major-gene-only models of BRCA1 and BRCA2, the former based on the study of highly selected families by Ford *et al.* (1998), the latter on a more recent study by Antoniou *et al.*

(2003). Moreover, Gui *et al.* used the same definition of family history as we have, namely two FDRs affected before age 50. Table 3.6 compares our premium rates with theirs, all as percentages of the standard premium. Although Gui *et al.* (2006) was based on a relatively unselected population, they assumed that the onset rates of BC and OC among BRCA1/2 mutation carriers were either 100% or 50% of the rates estimated, as a rough allowance for any remaining ascertainment bias; both are shown in the table.

Our full model (P + MG) yields slightly greater premiums, compared with Gui *et al.* if onset rates were 100% of those estimated. This is explained by the inclusion of the strong polygenic component. If we attribute all the inherited BC or OC cases to BRCA1/2 mutations we will estimate a higher frequency of such mutations in the population, and increase the probability of finding a major gene mutation carrier in a family with a history of BC or OC. By including the polygene we reduce the estimated frequency of BRCA1/2 mutations but, with the polygenic component acting in tandem with the BRCA1/2 mutations, the risk of BC is raised for any individual who has a family history.

Since there is no objective measure of how much the onset rates of Gui *et al.* or ours may have been affected by ascertainment bias, we tentatively conclude that our results show that the inclusion of polygenic inheritance (in addition to major gene inheritance) inflates the probable risk present in those that have developed a family history.

3.2.6 Genotype Distributions among those without a Family History

The histograms in Figures 3.1 and 3.2 show the genotype probability distributions among females with a family history. The corresponding histograms for females who do not have a family history are given in Figure 3.3, for ages 30 and 40, and Figure 3.4, for ages 50 and 60. We make three remarks on these:

(a) Relative to those with a family history, the probability of a family or applicant

carrying a major gene without presenting a family history is very small. In other words, an individual selected randomly from a population of females who have family histories is much more likely to have a BRCA1/2 genotype than an individual selected randomly from a population who do not have family histories.

(b) At ages 30 and 40 the polygenotype distribution among those without a family history is approximately the same distribution as that of the population at birth, *i.e.* binomial.

(c) At ages 50 and 60 the polygenotype distribution begins to show a slight positive skew. This is due to the difference in the rates of 'survival as Healthy' between high and low polygenotype carriers.

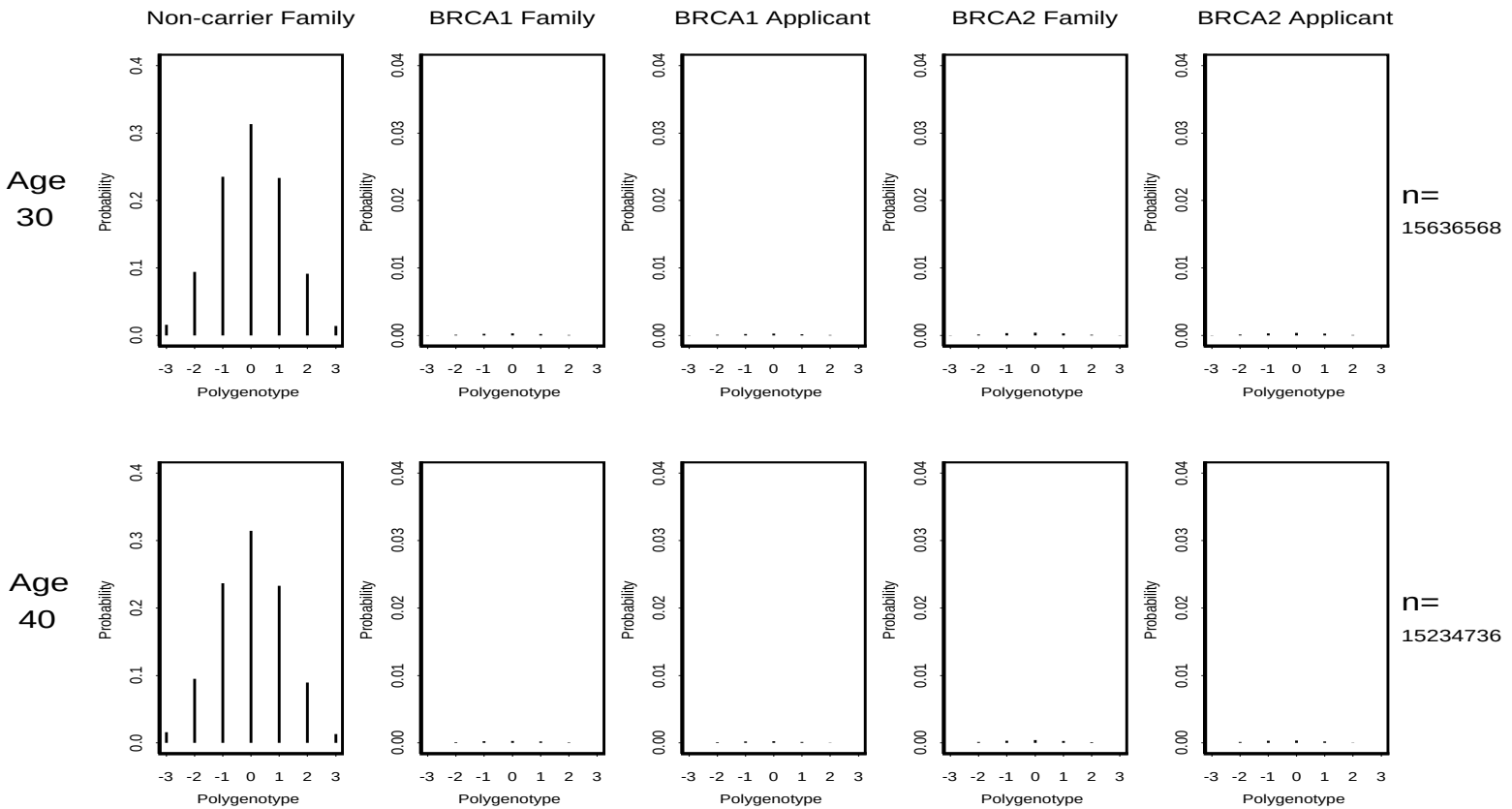


Figure 3.3: The distribution of polygenotypes by major genotype among healthy daughters aged 30 and 40, who do not have a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families.

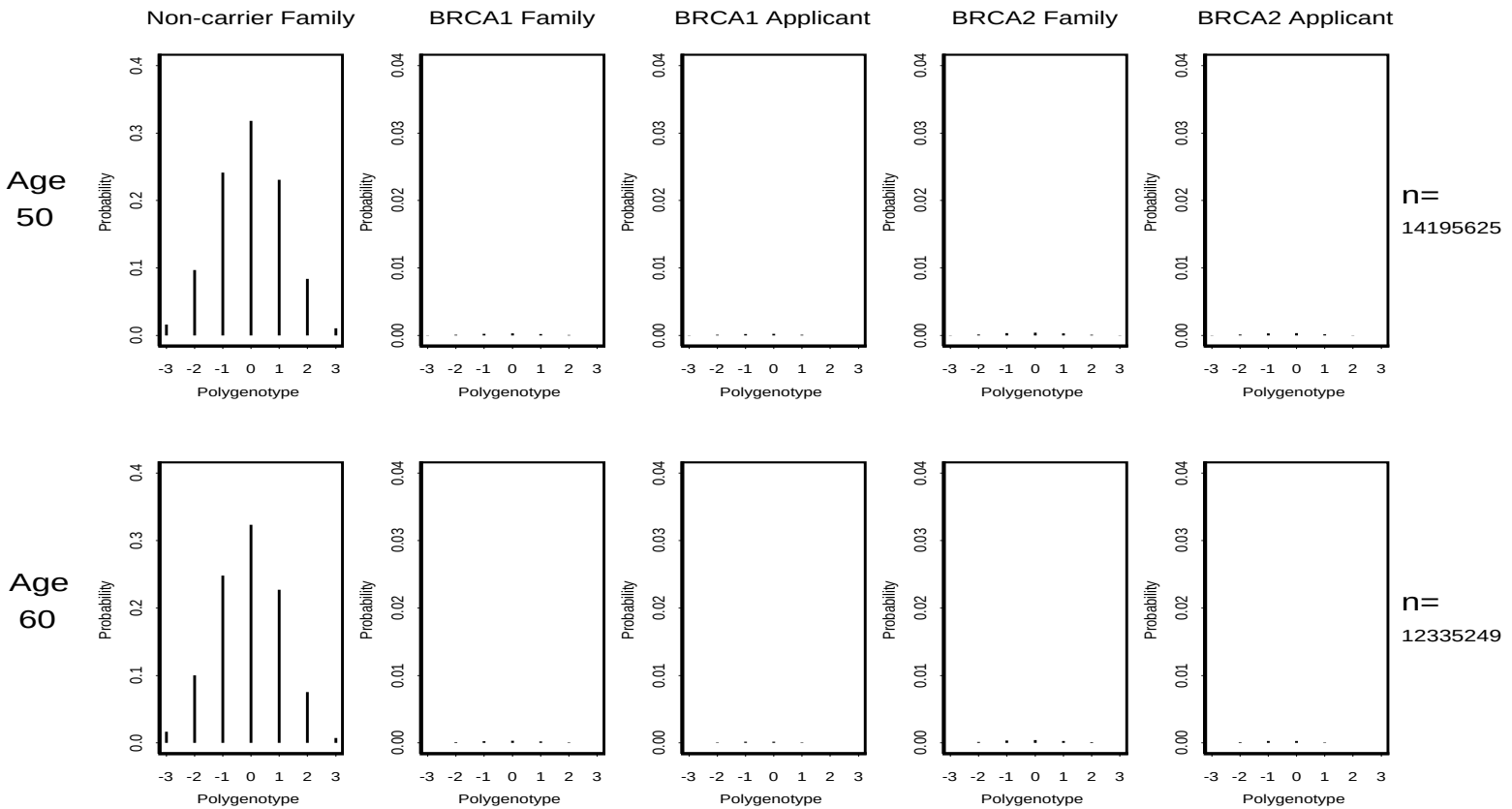


Figure 3.4: The distribution of polygenotypes by major genotype among healthy daughters aged 50 and 60, who do not have a family history. Based on 10,000,000 simulated families. The total number of individuals is shown on the right. Note different vertical scale for non-carrier families.

3.3 Conclusions

This is the first actuarial study to incorporate a fitted model of a polygenic disorder. The following conclusions might be relevant to GAIC when reviewing applications to use genetic test results for BC, or other polygenic diseases, in insurance underwriting.

Very substantial variation in premiums is attributable to the polygenic component of BC and OC risk, as opposed to the much-studied BRCA1 and BRCA2 major genes. Most significantly, some BRCA1/2 mutation carriers could be offered the standard premium rate after a genetic test that accounts for polygenotype. In the context of a lenient moratorium such as that in the UK (that is, a moratorium that allows insurers to use a genetic test result if it is to the applicant's advantage) this raises the possibility that a counteracting polygene configuration could be used to void a known BRCA1/2 mutation. At this stage, this is a brave extrapolation from a theoretical polygenic model, but enough genetic variation is unaccounted for by BRCA1 and BRCA2 to make such a conclusion plausible if and when polygenes become a therapeutic target for BC.

The polygenotype variation in the population (particularly, owing to its size, the subpopulation carrying no BRCA1/2 mutation) could raise questions that have so far largely been avoided because of the rarity of single-gene late-onset disorders. There appears to be enough variation in the risk attributed to the polygenotype that a test for an individual's polygenotype would raise new issues of adverse selection in the insurance market. This will be the subject of two of the chapters which follow.

However, our results are consistent with those of Macdonald *et al.* (2003b) and Gui *et al.* (2006) in showing that knowing of a BRCA1/2 mutation only (averaging over polygenotypes) presents a risk high enough to justify increased premiums, beyond the limits of any moratorium that may be in force. Although more recent epidemiology of BRCA1 and BRCA2 have suggested lower penetrance than originally estimated, the fact remains that BRCA1/2 mutation carriers are exposed to a much higher risk of BC and OC.

Because much of the genetic variation in BC can be explained by polygenes which affect the entire population (rather than just mutation carrier families), and the mode

of transmission is not Mendelian, it would seem that a woman with a family history need not have a polygenotype close to that of her sister. For example, parents with polygenotypes $(P_m, P_f) = (0, 0)$ can produce a child with polygenotype $P_c = +3$ with the same probability as they can produce a child with polygenotype $P_c = -3$. One sister at high risk of BC does not make it certain that any of her sisters will be also. This would partially explain why a polygenic component, which accounts for about 75% of familial risk of BC (approximately three times that of BRCA1/2 mutations), does not inflate family history premiums exceptionally. Thus when we use different models of inherited BC risk we find different premium ratings for a family history. We have also found a large difference in premium ratings if the definition of family history is tightened. Possibly ≥ 3 affected members rather than ≥ 2 affected members is the reasonable threshold of serious risk beyond which insurance may not be attainable.

Chapter 4

Estimating the Costs of Adverse Selection

4.1 Introduction

4.1.1 The UK Moratorium on Insurers' Use of Genetic Information

The current moratorium on insurers' use of genetic tests in the UK prevents insurers from accessing DNA-based genetic test results, but allows them access to the quasi-genetic information contained in a family history of disease. The moratorium introduces information asymmetry, whereby insurance applicants are more aware of their risk than is the insurer. Such applicants have the ability to 'adverse select' against the insurer: they may purchase more insurance coverage than they would were they charged the premium rate appropriate for their true risk. It is this possibility that makes insurers wary of a ban on using test results. Our goal is to investigate whether these fears are well-founded in the context of the new polygenic model for BC risk.

Let us consider an example. A woman undergoes a genetic test and discovers she carries a risky genotype. She then buys more insurance than she would have done without knowing the result. She has made two decisions: (a) to take the genetic test;

and (b) to buy a certain level of insurance in the light of the result. The first may be influenced by the existence of a screening program, possibly just for individuals with a family history, or by the availability of effective clinical interventions. The second is very difficult to research, and we often fall back on the assumption that the person tested will behave as a rational agent in an economist's model. It is possible that individuals might quite soon be able to access test results that list a large number of genetic variations. When this happens the difference between 'access' and 'interpret' will become interesting.

4.1.2 Major Genes and Polygenes

The link between high risks of BC and OC, and rare mutations in either of the BRCA1 and BRCA2 genes, is well-established. Several actuarial studies have considered the implications for the life and CI insurance markets (Gui *et al.*, 2005; Macdonald, Waters & Wekwete, 2003a, 2003b; Subramanian *et al.*, 2000; Lemaire *et al.*, 1999).

Adverse configurations of a polygene may confer susceptibility to a particular disease, or beneficial configurations might protect against it. These outcomes could also be strongly influenced by the environment. It is likely that we all carry some 'good' and some 'bad' polygene configurations, but this is quite speculative at this stage.

We use the estimated rates of onset of BC and OC in the model of Antoniou *et al.* (2002) which includes presence of the major genes BRCA1 and BRCA2, and a polygene affecting BC risk only. We describe this model in greater detail in Section 2.2.3. Recall that the polygenotype was assumed to act multiplicatively on the hazard rate of BC onset (see Equation 2.8) as follows:

$$\text{Hazard} = \text{Baseline Hazard} \times \exp(c \times \text{Polygenotype}), \tag{4.1}$$

where the baseline hazard is that for Polygenotype = 0, and the constant c is just a scale factor. Assuming each allele to be equally common, and inherited independently

of the others, the distribution in the population of the quantity (Polygenotype + 3) is Binomial(6,1/2).

What is significant is that we no longer consider only rare major genes, but also polygenes that are present (in a modest variety of configurations) in every individual. So, instead of tiny numbers of people with very high risks selecting against the insurer, larger numbers with modestly increased risks may contribute to the cost of adverse selection.

The goal of this chapter is to assess the possible impact of the polygene upon a hypothetical CI market. We do this by constructing a multi-state model representing an individual's decision-making in terms of insurance and genetic testing. This allows us to estimate the costs of adverse selection under fixed assumptions about policyholders' behaviour.

4.2 Modelling a CI Insurance Market

4.2.1 Model Setup

To model the potential costs of adverse selection in a CI market we use the model in Figure 4.1. Each genotype is represented by a version of this model, with different rates of onset of BC and OC that correspond to the genotype. The model represents the life history of a person, as yet uninsured, who may buy insurance before or after having a genetic test. Level premiums are payable while in either of the insured states, and the benefits are payable on transition from either of these states into a 'critical illness' state (which represents the onset of BC, OC, or another critical illness).

Usually, determining the premiums payable while insured requires that the cash-flows depend on the age that insurance was purchased, hence they are duration-dependent and not Markov. Instead, we use a 'current-cost' basis for charging premiums where the premium payable at time t is equal to the expected claims that arise at time $t + dt$. In other words, the premium is a weighted average of all the intensities out of the healthy states (*i.e.* not the 'Critical Illness' or 'Dead' states) and into the 'Critical Illness' state, the weights being the occupancy probabilities in

each of the healthy states. This rate of premium is independent of the time of entry into an 'Insured' state. We assume that all policies that are purchased in our market expire when the insured reaches age 60.

Some of the assumptions we make are as follows:

(a) Large and small markets are represented by insurance purchase at a 'normal' rate of 0.05 or 0.01 *per annum*, respectively.

(b) In both markets, low risk polygenotype carriers may buy less insurance than the 'normal' rate. These carriers purchase at the same rate as the normal rate, half of the normal rate, or at rate zero.

(c) Genetic testing occurs at three possible annual rates: 0.02972 (low), 0.04458 (medium), or 0.08916 *per annum* (high), based on an uptake proportion of 59% (Ropka *et al.*, 2006) over a period of 30, 20, or 10 years of testing respectively. Also, testing may only occur between ages 20 and 40 (when testing has high priority).

(d) 'Severe' adverse selection means that high-risk polygenotype carriers will purchase insurance at rate 0.25 *per annum*.

All other intensities, governing transitions into the 'Dead' and 'Critical Illness' states, are as were used in the CI pricing model in Figure 2.3.

EPVs of benefits and premiums are found by solving Thiele's differential equation backwards numerically with force of interest $\delta = 0.05$. Occupancy probabilities are found by solving Kolmogorov's Forward Equations. We described these methods in Section 1.4. In both cases we use a Runge-Kutta algorithm with a step-size of 0.0005.

4.2.2 A Genetic Screening Program for the Polygene Only

For simplicity, the first possibility we consider is that a genetic screening program exists for the polygenotype only, not extending to the BRCA1/2 genotypes. There are seven polygenotypes, therefore a 42-state model. We assume that the distribution of new-born persons in the seven sub-populations is Binomial(6,1/2), and that mortality

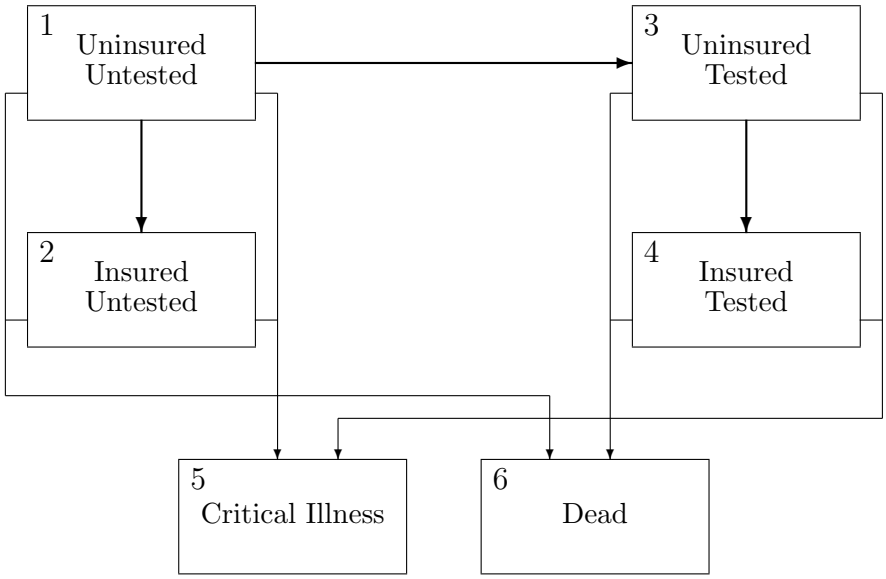


Figure 4.1: A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing is available at an equal rate to all subpopulations.

and morbidity before age 20 does not depend on genotype (so that the expected proportions in each starting state at age 20 have not changed). Since the rate of BC onset is negligible before about age 30, this assumption seems reasonable.

Tested carriers may alter their insurance-buying habits, in one of two ways: carriers of deleterious polygenotypes may buy more insurance, or carriers of protective polygenotypes may buy less insurance. This latter behaviour is uncommon in adverse selection studies; it is usually assumed that individuals who receive negative test results will purchase insurance at the normal market rate. One study (Subramanian *et al.*, 1999) performed a sensitivity analysis where tested non-carriers could reduce their coverage. It is plausible that this makes more sense from an economic point of view. Figure 4.2 shows three scenarios of differing severity.

The percentage by which all premiums must be raised in order to negate the adverse selection costs is given by:

$$100 \times \left(\frac{\text{EPV[Loss|Adverse Selection]} - \text{EPV[Loss|No Adverse Selection]}}{\text{EPV[Premium Income|Adverse Selection]}} \right). \quad (4.2)$$

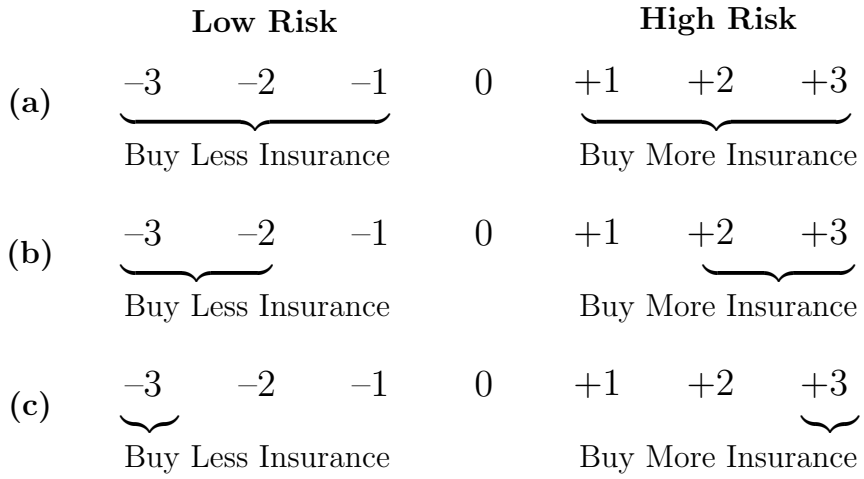


Figure 4.2: Three possible behaviours of tested polygenotype carriers in the adverse selection model, labelled (a), (b) and (c).

We will refer to these percentages as the 'costs of adverse selection'. The expected present value of premium income, given adverse selection occurs, is simply:

$$\text{EPV[Benefits|Adverse Selection]} - \text{EPV[Loss|Adverse Selection]}. \tag{4.3}$$

When there is no adverse selection the expected present value of losses is zero.

Table 4.1 shows the premium increases needed to absorb the costs of the severe adverse selection under each scenario (a), (b) or (c). Compared with previous results based on major genes only (Gui *et al.*, 2006; Gutiérrez & Macdonald, 2007) these are very high. This is because deleterious polygenotypes are relatively more common, and also because these authors did not consider the possibility that carriers of beneficial genotypes would buy less insurance (since 'beneficial' in their studies simply meant 'normal'). Note the large fall in costs between Scenarios (b) and (c). This is explained by the small size of the adverse-selecting groups in Scenario (c), essentially the extreme tails of the Binomial distribution of polygenotypes.

Curiously, in a small market the cost of adverse selection is always higher in Scenario (b) than in (a). This is because premium increases are relative to a baseline 'ordinary' rate (OR rate), which is higher in Scenario (a) than in (b).

We repeated the experiment for the case where high risk polygenotype carriers

Table 4.1: Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Screening available for the polygene only.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	1.05975	0.90051	0.26447
		Half	1.69947	1.42748	0.30825
		Nil	2.86994	2.36206	0.38349
	Small	Normal	6.81382	7.03421	1.03701
		Half	7.72964	7.90872	1.10420
		Nil	8.86328	8.94001	1.17995
Medium	Large	Normal	1.39994	1.20315	0.29909
		Half	2.27895	1.93093	0.35897
		Nil	3.95151	3.25381	0.46294
	Small	Normal	8.25781	9.04200	1.32143
		Half	9.49144	10.27268	1.41322
		Nil	11.10982	11.75999	1.51728
High	Large	Normal	2.01615	1.77037	0.36453
		Half	3.40261	2.92799	0.45793
		Nil	6.32401	5.15661	0.62418
	Small	Normal	10.31240	12.36857	1.83292
		Half	12.19959	14.37086	1.97494
		Nil	15.09918	16.92565	2.13791

purchase insurance at the normal rate, hence they do not adverse select. However low risk polygenotype carriers may still modify their purchasing behaviour by purchasing at half the normal rate or not purchasing at all. Table 4.2 shows the effect of the low risk polygenotype carriers behaviour upon the cost of adverse selection. When the low risk polygenotype carriers purchase at the normal rate, all subpopulations in the model are purchasing at the normal rate and so there is no adverse selection.

4.2.3 A Genetic Screening Program for the Polygene and Major Genes

Now we continue with the same testing scenario as depicted in Figure 4.1 but consider the possibility that screening is available for the major BRCA1 and BRCA2 genotypes, as well as the polygenotype. Thus we have $3 \times 7 = 21$ distinct genotypes, hence subpopulations, and 126 states in the model. Using Equation (2.11), the carrier frequencies of BRCA1 and BRCA2 mutations are estimated at 0.0010181 and 0.0013577 respectively (Antoniou *et al.*, 2002), hence we fix the proportions in the relevant subpopulations.

We consider the same adverse selection scenarios as in Figure 4.2, but additionally those who carry adverse BRCA1/2 mutations select against the insurer regardless of polygenotype. The resulting premium increases are shown in Table 4.3. They are not much larger than those in Table 4.1, the greatest increase being in Scenario (c). Compared with screening for the polygene alone, the adverse selection costs borne by insurers if screening is extended to BRCA1/2 mutations are not high.

We also considered the unusual possibility that some BRCA1/2 mutation carriers do not adverse select because they carry protective polygenotypes which 'void' the BRCA1/2 risk. This was shown in Section 2.3.2, when it was apparent that BRCA1/2 females carrying polygenotype -3 could plausibly obtain CI insurance at ordinary rates. When we took account for this the change to the results was minor, and hence we omit them.

Table 4.2: Costs of adverse selection resulting from low risk polygenotype carriers buying less insurance than normal in a critical illness insurance market open to females between ages 20–60. High risk polygenotype carriers buy insurance at normal rate. Screening available for the polygene only.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	0.00000	0.00000	0.00000
		Half	0.64468	0.51917	0.16535
		Nil	1.82361	1.44036	0.44408
	Small	Normal	0.00000	0.00000	0.00000
		Half	0.98952	0.78697	0.24367
		Nil	2.19323	1.71326	0.51771
Medium	Large	Normal	0.00000	0.00000	0.00000
		Half	0.88799	0.71343	0.22642
		Nil	2.57479	2.01092	0.61215
	Small	Normal	0.00000	0.00000	0.00000
		Half	1.36056	1.07732	0.33161
		Nil	3.08502	2.37443	0.70689
High	Large	Normal	0.00000	0.00000	0.00000
		Half	1.40648	1.12406	0.35433
		Nil	4.35040	3.28932	0.97370
	Small	Normal	0.00000	0.00000	0.00000
		Half	2.13881	1.67679	0.51027
		Nil	5.14799	3.79710	1.09563

Table 4.3: Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Screening available for major genes and the polygene.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	1.08112	0.93445	0.34798
		Half	1.73147	1.47837	0.53487
		Nil	2.92044	2.44180	0.84843
	Small	Normal	6.92408	7.25457	2.86768
		Half	7.85526	8.15781	3.15438
		Nil	9.00878	9.22405	3.47696
Medium	Large	Normal	1.42838	1.24882	0.46798
		Half	2.32238	2.00052	0.72474
		Nil	4.02274	3.36579	1.16023
	Small	Normal	8.39054	9.32273	3.81134
		Half	9.64599	10.59581	4.20875
		Nil	11.29504	12.13682	4.65893
High	Large	Normal	2.05799	1.83889	0.69839
		Half	3.46956	3.03642	1.10290
		Nil	6.44544	5.34242	1.80734
	Small	Normal	10.47855	12.75016	5.53492
		Half	12.40261	14.82749	6.16727
		Nil	15.36631	17.48591	6.89426

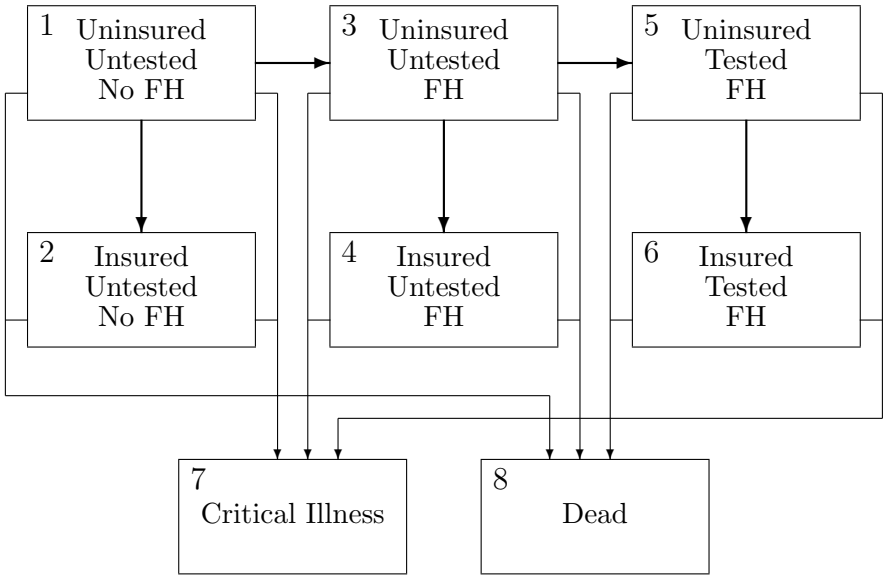


Figure 4.3: A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing is available only after the appearance of a family history (FH) of BC/OC.

4.2.4 More Limited Genetic Testing for the Polygene and Major Genes

Only about one quarter of BC cases are hereditary and only about one fifth of these are caused by identifiable (major) genes, so mass screening programs are mostly ineffective. Much more likely is that testing is offered only to women who present a family history of BC/OC, who are more likely to carry deleterious genes. Therefore we adjust our original model as shown in Figure 4.3, to include the development of a family history, which in turn is a prerequisite for genetic testing.

The rate of genetic testing among those individuals who have developed a family history we take as a constant intensity amounting to a proportion of 70% (Ropka *et al.*, 2006) becoming tested over a 30, 20 or 10 year period, representing low, medium and high levels of genetic testing, respectively.

We define a family history to mean a healthy woman has two or more first-degree relatives who contracted BC/OC before age 50. The incidence of family history was

calculated using the formula common in epidemiology:

$$\text{Incidence Rate} = \frac{\text{Number of new cases arising in specified time period}}{\text{Number of individuals at risk during the time period}}. \qquad (4.4)$$

In order to be 'at risk' of a family history developing a daughter must be healthy with either no others affected by BC or OC and at least two other healthy females under the age of 50, or one other female affected with BC or OC under the age of 50 and at least one other female healthy and under 50. If a family history develops, each healthy daughter contributes as a 'new case' in the incidence at that time. These rates were estimated by the simulation in Section 3.2.1. The incidence rate is shown in Figure 4.4 for the subpopulations without BRCA mutations in the family, and in Figure 4.5 for the BRCA1 and BRCA2 carrier families. Since all siblings are assumed to be the same age, after age 50 the family history threshold cannot be crossed and the incidence rate is zero.

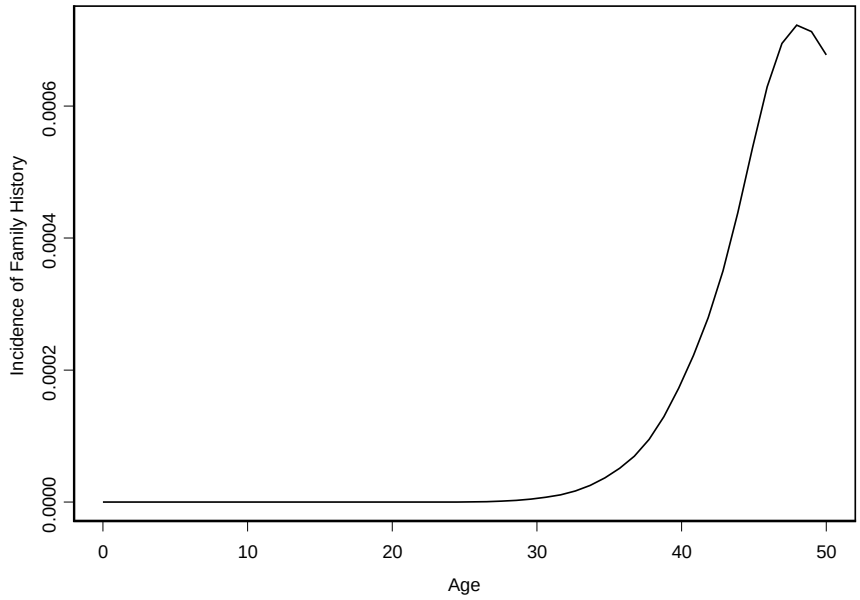


Figure 4.4: The incidence of family history for the subpopulations without BRCA mutations. A family history may not appear beyond age 50 in any subpopulation.

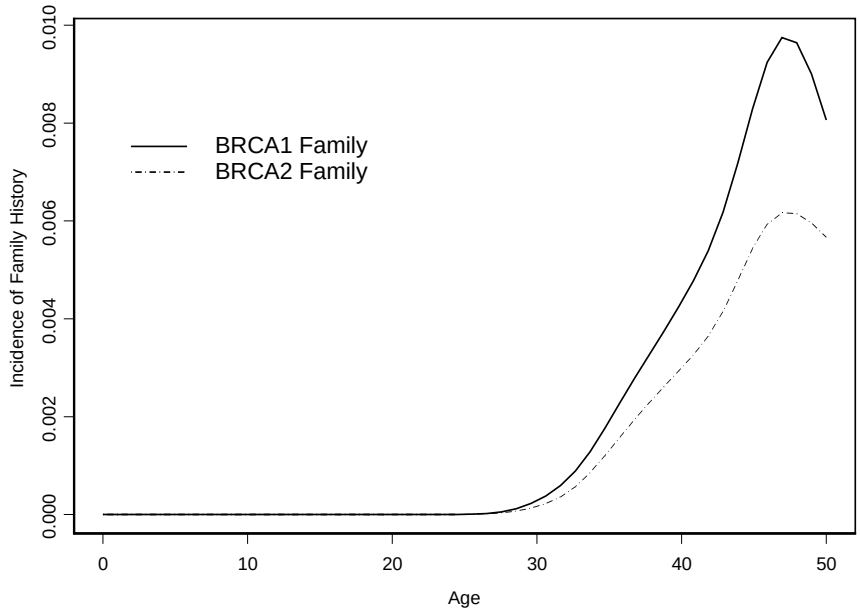


Figure 4.5: The incidence of family history for the subpopulations with BRCA1/2 mutations in the family. A family history may not appear beyond age 50 in any subpopulation.

We give our results in Table 4.4. The costs of adverse selection are greatly reduced when a family history is a prerequisite for a genetic test. Once again a small insurance market suffers higher relative costs.

As in Section 4.2.3, we also considered the possibility that some BRCA1/2 mutation carriers do not 'adverse select' because they carry protective polygenotypes which 'void' the BRCA1/2 risk. This had a negligible effect, and we omit the results. To some extent this outcome strengthens the case for offering CI insurance at ordinary rates to such individuals.

4.2.5 Separate Testing for Polygene and Major Genes

We can imagine a more realistic situation where testing for major genes is conducted through a public health service, once a family history has rendered the patient 'at risk', and testing for the polygene may be sought privately (by asymptomatic individuals).

Table 4.4: Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Testing available for major genes and the polygene after the onset of a family history.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	0.00026	0.00025	0.00020
		Half	0.00034	0.00031	0.00025
		Nil	0.00044	0.00041	0.00031
	Small	Normal	0.00180	0.00164	0.00118
		Half	0.00192	0.00175	0.00125
		Nil	0.00206	0.00187	0.00134
Medium	Large	Normal	0.00034	0.00032	0.00025
		Half	0.00044	0.00041	0.00032
		Nil	0.00059	0.00055	0.00041
	Small	Normal	0.00255	0.00232	0.00165
		Half	0.00273	0.00248	0.00177
		Nil	0.00292	0.00266	0.00189
High	Large	Normal	0.00053	0.00049	0.00038
		Half	0.00071	0.00066	0.00049
		Nil	0.00098	0.00089	0.00065
	Small	Normal	0.00445	0.00404	0.00287
		Half	0.00477	0.00433	0.00306
		Nil	0.00511	0.00463	0.00327

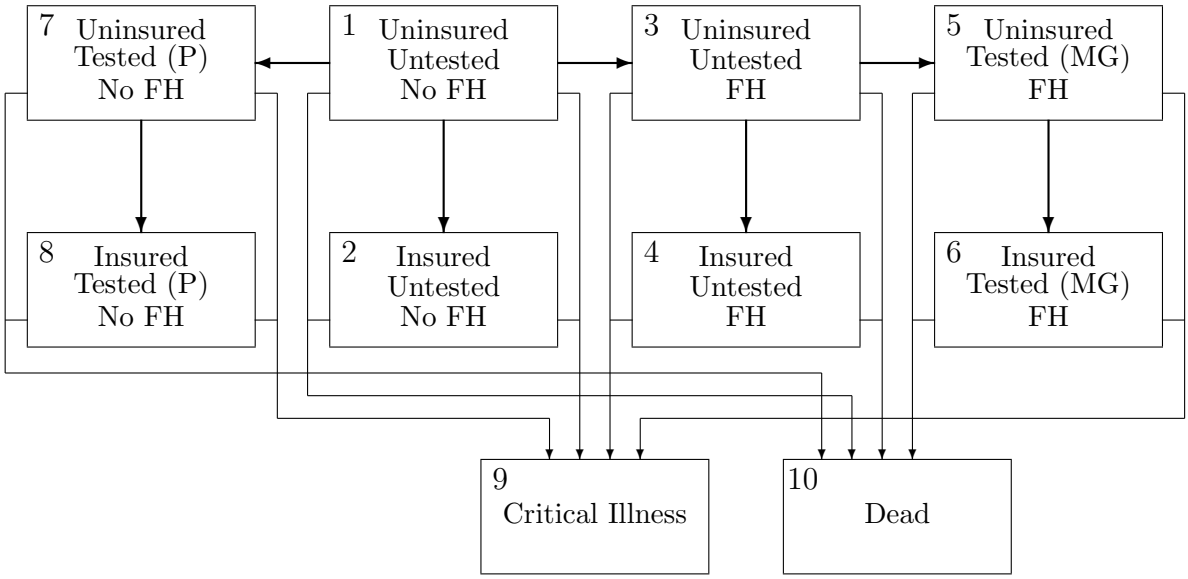


Figure 4.6: A model of the behaviour of a genetic subpopulation with respect to purchasing of CI insurance. Genetic testing for major genes (MG) is available only after the appearance of a family history (FH) of BC/OC. Testing for the polygene (P) is available before a family history has appeared.

Therefore, we have two different testing events: one for the BRCA1/2 genes and one for the polygene. Thus we model each subpopulation using the state-space shown in Figure 4.6. Both the family-history related and the non-family-history related testing rates may be at the low, medium or high levels.

Our results are shown in Table 4.5. It is interesting to compare this with Table 4.3. Throughout, the 'separate testing' model has somewhat smaller costs than the 'combined testing' model. In fact, the costs are close to those of the polygene-only screening program (Table 4.1).

We have previously assumed that 'severe' adverse selection takes place, and that this is characterised by a rate of insurance purchase of 0.25 *per annum* by high risk polygenotype carriers and BRCA1/2 mutation carriers. We assume a less severe rate of adverse selection to be 0.1 *per annum*. Table 4.6 gives the costs of adverse selection when we apply this more modest rate of adverse selection. Note that the costs are still very high in the small market and in general there has only been a small reduction throughout all the costs.

Table 4.5: Costs of severe adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Separate testing for polygene and major genes.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	1.04241	0.89201	0.30311
		Half	1.67193	1.41431	0.46840
		Nil	2.82281	2.34075	0.74706
	Small	Normal	6.71995	6.97657	2.53289
		Half	7.61851	7.84459	2.78778
		Nil	8.72852	8.86823	3.07451
Medium	Large	Normal	1.37917	1.19288	0.40759
		Half	2.24492	1.91474	0.63453
		Nil	3.88979	3.22682	1.02114
	Small	Normal	8.15742	8.97684	3.36875
		Half	9.36749	10.19844	3.72104
		Nil	10.95022	11.67450	4.11975
High	Large	Normal	1.99302	1.75879	0.60791
		Half	3.36192	2.90895	0.96494
		Nil	6.23846	5.12262	1.58888
	Small	Normal	10.22137	12.30461	4.89643
		Half	12.07653	14.29422	5.45403
		Nil	14.91496	16.83171	6.09361

Table 4.6: Costs of modest adverse selection resulting from high risk polygenotype carriers buying more insurance than low risk polygenotype carriers in a critical illness insurance market open to females between ages 20–60. Separate testing for polygene and major genes.

Genetic Testing	Market Size	Insurance Purchasing of Low Risk Polygenotypes	Scenario (a) %	(b) %	(c) %
Low	Large	Normal	0.60275	0.50524	0.16798
		Half	1.23472	1.02422	0.33281
		Nil	2.39027	1.94512	0.61072
	Small	Normal	5.59863	5.54929	1.94805
		Half	6.51345	6.40019	2.20002
		Nil	7.64109	7.40364	2.48350
Medium	Large	Normal	0.80890	0.68244	0.22773
		Half	1.67911	1.39829	0.45385
		Nil	3.33224	2.70007	0.83911
	Small	Normal	6.92804	7.23268	2.61005
		Half	8.16462	8.42626	2.95730
		Nil	9.77516	9.86818	3.35034
High	Large	Normal	1.20769	1.03200	0.34723
		Half	2.58640	2.16860	0.70242
		Nil	5.47923	4.35764	1.32324
	Small	Normal	8.96365	10.18049	3.86596
		Half	10.86549	12.11832	4.41319
		Nil	13.74869	14.58806	5.04095

4.3 Conclusions

The key feature of this work is the move from looking purely at rare single genes to looking at combinations of genes that are common in the population. The consequence of this is that adverse selection becomes an option for a larger proportion of the population, hence the potential for adverse selection is much greater. Furthermore, the relative risks attributed to the polygene, as estimated by Antoniou *et al.* (2002), are as high as 23.62 and as low as 0.04; many times more extreme than the assumptions in Macdonald, Pritchard & Tapadar (2006) for example.

This work highlights the inherent risk of a moratorium in the presence of genetic screening. When we look at the most adverse purchasing behaviour, Scenario (a), in a small insurance market the necessary premium increase approaches 17% (under the assumption of a high level of genetic testing). It is difficult to state what level of costs would seriously burden the insurer, but 17% is certainly not likely to go unnoticed. We must also bear in mind that this cost arises solely from the contribution of BC/OC, and that there are several other genetic disorders (the ABI has a list of seven) which will certainly increase this estimate.

The inflated costs in the small market are perhaps a cause for concern to any emerging CI markets. Should genetic technology advance to a level where polygene testing becomes available, youthful insurance markets are exposed to high costs from a moratorium on genetic testing (even at modest levels of testing). This should be a consideration when seeking to restrict insurers' access to information.

Genetic testing following family history onset presents lower costs to the insurance industry than a genetic screening program. Such a testing procedure is the more realistic and cost-effective method but it is not uncommon (59%) for individuals to submit to testing if they have no familial risk (Ropka *et al.*, 2006). The two, however, are not mutually exclusive (we have modelled a hybrid of the two cases).

We assumed that the behaviour of individuals aware of their polygenotype is dictated by one of three possibilities (Figure 4.2). These are our best estimates of an individual's behaviour in the same sense that the polygenic model is the best estimate of polygene risk; a compromise of realism and simplicity. There is great variability

in the costs resulting from the different insurance-buying behaviours (about 10-fold in the large market with a medium level of testing). By introducing new genetic discoveries into the adverse selection model we have inevitably added to the list of significant but uncertain parameters that we require.

Chapter 5

Estimating the Extent of Adverse Selection

5.1 Introduction

5.1.1 A Review of Economic Modelling of Adverse Selection

One key feature of the multiple-state model approach to estimating the costs of adverse selection is that it is assumed that individuals who obtain an adverse genetic test result will be highly likely to buy insurance within a short time-frame. For example, review the rather simplistic assumptions (see Figure 4.2) that we made in the previous chapter. This lacked sound economic rationale. In the economics literature several papers have considered the effects of information asymmetry in insurance markets.

Doherty & Thistle (1996) discuss the economic value of information held by potential insurance applicants. They highlight that under symmetric information (where insurers are informed of all test results) risk-averse individuals would be deterred from obtaining diagnostic genetic tests. But with test results unavailable to the insurer, an individual would desire test information to enable an informed decision on insurance purchasing. They concluded that any loss of efficiency in the market should be weighed against the value of private information.

Hoy & Witt (2005) consider the specific case of BRCA1/2 genetic test results and

life insurance. They simulated the market for ten-year term assurances offered to women aged 35 to 39. They warn that the efficiency of the market may be compromised if a sufficiently large fraction of women undertake genetic testing. They suspect this concern to be greater when a wide range of genetic tests, for many different disorders, become available to the public.

In respect of multifactorial disorders, 'adverse' genotypes confer relatively modest excess risk to a particular disease. In this case, heavy-handed assumptions of insurance purchasing behaviour are arguably ill-suited. Macdonald & Tapadar (2006) used a model of gene-environment interaction to determine the likelihood of adverse selection. They found no convincing evidence that privacy of information would cause adverse selection to be a serious risk.

In this chapter we shall apply some of the principles of utility theory to assess how a population that carries the BC polygene may effect adverse selection.

5.2 Utility Models

5.2.1 Utility Functions

We will use utility functions to describe the motivation for an individual to insure. The utility function $U(w)$ can be interpreted as an increasing concave relation between wealth W and the relative satisfaction gained from holding wealth W. Macdonald & Tapadar (2006) parameterised four utility functions, three from the Iso-Elastic family of utility functions and one from the Negative Exponential family. Throughout this work we shall use the same utility functions, shown in Table 5.1. All of these functions satisfy $U'(w) > 0$ and $U''(w) < 0$, hence they are concave increasing. In the rest of the paper they will be referred to as Model I, Model II, Model III and Model IV, as shown in Table 5.1. Models I and II have low risk aversion whereas models III and IV were parameterised using data from a 1995 Italian thought experiment (Eisenhauer & Ventura, 2003), adjusted for the sterling/lira exchange rate and UK price inflation up to 2006, and have higher risk aversion.

Table 5.1: The four utility functions parameterised by Macdonald & Tapadar (2006).

Family	Utility Function	Parameter	Model
Iso-Elastic	$U(w) = \begin{cases} (w^\lambda - 1)/\lambda & \lambda < 1 \text{ and } \lambda \neq 0 \\ \log(w) & \lambda = 0 \end{cases}$	$\lambda = 0.5$	I
		$\lambda = 0$	II
		$\lambda = -8$	III
Negative Exponential	$U(w) = -\exp(-Aw)$	$A = 9 \times 10^{-5}$	IV

The shape of each of these utility functions is displayed in Figure 5.1. The graphs help to show how we model the utility that an individual receives from holding different amounts of wealth. As a rough guide: the greater the curvature of $U(w)$ the higher the risk aversion.

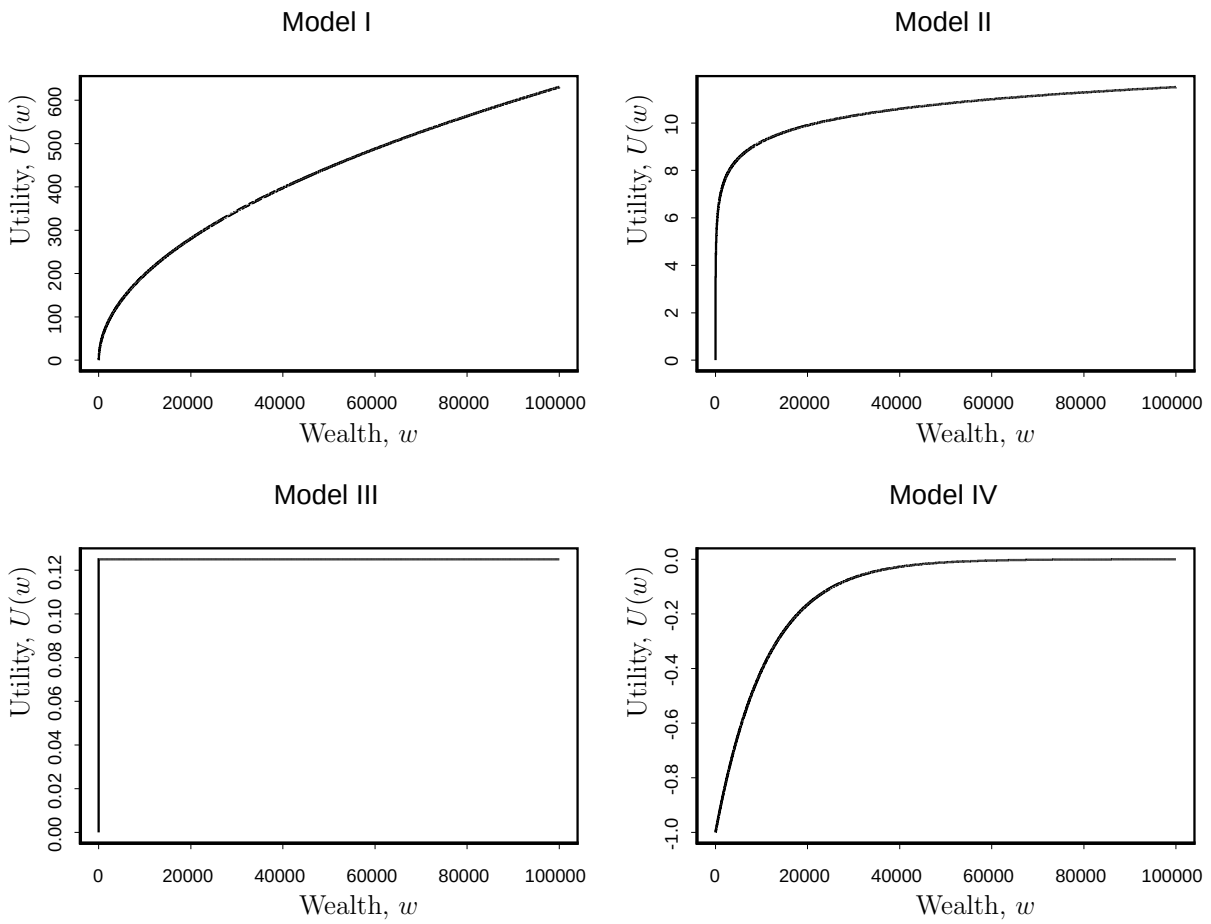


Figure 5.1: The four utility models given in Table 5.1 for wealth, w, between 0 and 100,000 pounds.

We assume that in the face of uncertainty a normal risk-averse individual will seek to maximise his or her expected utility. For example, suppose an individual with total wealth W faces a loss L with probability q. The *actuarial value* (or *fair value*) of insurance against this random event is qL, but the individual would be prepared to pay premium $X \geq qL$ as long as:

$$U(W - X) > qU(W - L) + (1 - q)U(W). \tag{5.1}$$

The premium X at which an individual would stop purchasing is found by converting the inequality in Equation (5.1) to an equality and solving for X.

Now let us imagine that it is possible to stratify the population into separate subpopulations each with a different probability of encountering a loss. Consider that it is also possible that individuals are able to discover their personal risk (perhaps by private consultation or testing) and can deduce their own probability of a CI. Those who have a perceived probability that is less than X/L will not buy into insurance at that price and the higher-risk individuals who find the price acceptable do buy insurance. This is adverse selection.

5.2.2 Notation for the Polygenic Model

Antoniou *et al.* (2002) (see Section 2.2) were the first to take the hypothetical polygene model for BC risk and fit it to a large population. Recall that the intensities of onset of BC are given as:

$$\mu^{BC}(x, R) = \mu^{BC}(x)\ RR_M\ e^R, \qquad \text{where } R \approx \frac{P}{\sqrt{n/2}}\ \sigma_R, \tag{5.2}$$

and RR_M is the relative risk of carrying a mutated BRCA1/2 gene. The polygenotype P is an integer between $-n$ and n, representing all possible distinct configurations of alleles in the polygene. In the model of Antoniou *et al.* (2002) $n = 3$ and σ_R was estimated as 1.291.

For this chapter alone, we assume that the proportions of the population that carry a given polygenotype or major genotype remain fixed from birth for all ages

(where before the proportions changed because of differences in the rate of morbidity between subpopulations). Therefore it is assumed that the probability of selecting any polygenotype from the population is binomial with parameters $(2n, 1/2)$. Hence, the proportion of individuals with polygenotype p and major genotype m is $\omega(p, m) = \omega_P(p)\omega_M(m)$ where:

$$\omega_P(p) = \begin{pmatrix} 2n \\ p+n \end{pmatrix} \left(\frac{1}{2}\right)^{2n} \text{ and } \omega_M(m) = \begin{cases} 1 - p_1 - p_2 & \text{for} \quad m = 0 \\ p_1 & \text{for} \quad m = 1 \\ p_2 & \text{for} \quad m = 2 \end{cases} \tag{5.3}$$

and where p_1 and p_2 are the genotype carrier probabilities of the BRCA1 and BRCA2 genes respectively (given in Section 4.2.3).

5.3 The Purchase of Critical Illness Insurance

5.3.1 Critical Illness Premiums

Wealth may be radically reduced by the incidence of a critical illness and so the concept of utility can be used to determine how much wealth an individual is willing to forego to obtain indemnity. If low risk individuals are not prepared to pay for insurance priced at the insurers' fair value there is adverse selection and we are left with generally higher risks in the insurance pool.

We will use the CI model in Figure 2.3 to calculate single premiums for CI policies. A unit sum assured is payable on transition from the Healthy state to any CI state (BC, OC or Other Critical Illness). For simplicity and consistency with previous studies of insurance and utility (Hoy & Witt, 2005; Macdonald & Tapadar, 2006) let the force of interest $\delta = 0$. This means that EPVs are equivalent to the probabilities of the CI event occurring.

Let $X(p, m)$ represent the single premium charged to individuals with polygenotype p and major genotype BRCAm. Table 5.2 shows the premiums for individuals with polygenotype -3 and -2 and no BRCA mutation. Since polygenotype -3 is

more protective against BC than polygenotype -2 the premium is smaller for the $P = -3$ population.

Table 5.2: Single premiums for various term assurances for the $P = -3$ and $P = -2$ non-BRCA mutation carrier ($M = 0$) subpopulations.

		Premium	
Age	Term	$X(-3, 0)$	$X(-2, 0)$
20	10	0.005827	0.005859
	20	0.021664	0.021991
	30	0.063343	0.064562
	40	0.147844	0.150075
30	10	0.015972	0.016270
	20	0.058006	0.059204
	30	0.143225	0.145448
40	10	0.042863	0.043794
	20	0.129762	0.131764
50	10	0.091393	0.092609

5.3.2 Threshold Premiums

Let $\bar{X}$ be the single premium offered to all females when the insurer has no information regarding genotype status. Thus $\bar{X}$ is the premium per unit loss averaged over all genotypes:

$$\bar{X} = \sum_m \sum_p \omega(p, m) X(p, m), \tag{5.4}$$

which is the fair value that an insurer would offer given no extra information regarding genotype. Insurance will be bought as long as:

$$U(W - \bar{X}L) > XU(W - L) + (1 - X)U(W), \tag{5.5}$$

where X is the probability of a loss occurring. Given that we can calculate $\bar{X}$, and that W and L are quantities which we can fix, we wish to find X^*, the value of X that solves Equation (5.5) when the inequality is replaced by equality. Then X^* is the threshold premium for the onset of adverse selection. Table 5.3 and Table 5.4 show the values of X^* for a selection of CI policies and a range of losses L when initial wealth $W = £100,000$. We can see that as the ratio of loss to wealth increases the

threshold premium decreases, implying that the threat of near catastrophic losses will induce individuals to pay the premium $\bar{X}$ even if the premium for their personal risk is much lower than this.

In Table 5.2 we presented the premiums that would be charged to the $(-3, 0)$ subpopulation if genetic information were available to distinguish risk classes. Adverse selection will occur when $X(-3, 0) < X^*$, so in Table 5.3 and Table 5.4 we can see for what values of L this is the case and can roughly deduce what level of desired loss coverage will initiate adverse selection. These levels of loss (as a proportion of wealth) are generally around 0.85 for Model I and around 0.65 for Model II. However for models III and IV, $X^* < X(-3, 0)$ for almost all levels of loss we have tabulated. By setting $X^* = X(-3, 0)$ and solving the equality corresponding to Equation (5.5), we can find the levels of loss that would initiate adverse selection. These are given in Table 5.5. In Model I the loss that would initiate adverse selection is often equal to £100,000, in which case adverse selection will occur regardless of any value of the possible loss. We can see that adverse selection could occur under models III and IV, but only for very low levels of insured loss (for which CI insurance is certainly unnecessary).

Table 5.3: Premium rates X^* that are the thresholds at which adverse selection will take place, for a variety of CI policies and initial wealth $W = £100,000$.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	20	10	0.00653	0.00635	0.00616	0.00595	0.00572	0.00547	0.00519	0.00485	0.00442
		20	0.03004	0.02923	0.02836	0.02742	0.02640	0.02527	0.02397	0.02243	0.02042
		30	0.09326	0.09089	0.08833	0.08555	0.08250	0.07909	0.07518	0.07047	0.06426
		40	0.19895	0.19442	0.18949	0.18407	0.17804	0.17120	0.16324	0.15351	0.14044
	30	10	0.02375	0.02310	0.02241	0.02167	0.02085	0.01995	0.01893	0.01771	0.01612
		20	0.08759	0.08534	0.08293	0.08031	0.07744	0.07422	0.07053	0.06611	0.06027
		30	0.19428	0.18983	0.18499	0.17968	0.17377	0.16707	0.15927	0.14976	0.13699
	40	10	0.06620	0.06446	0.06261	0.06059	0.05839	0.05594	0.05313	0.04976	0.04534
		20	0.17627	0.17215	0.16768	0.16278	0.15734	0.15120	0.14406	0.13538	0.12376
	50	10	0.12052	0.11753	0.11431	0.11080	0.10693	0.10259	0.09758	0.09155	0.08354
Model II	20	10	0.00636	0.00601	0.00564	0.00525	0.00484	0.00440	0.00390	0.00334	0.00263
		20	0.02929	0.02770	0.02604	0.02428	0.02240	0.02037	0.01811	0.01551	0.01221
		30	0.09107	0.08642	0.08149	0.07624	0.07058	0.06439	0.05746	0.04938	0.03903
		40	0.19480	0.18588	0.17630	0.16591	0.15453	0.14186	0.12740	0.11020	0.08769
	30	10	0.02315	0.02188	0.02056	0.01917	0.01768	0.01607	0.01428	0.01223	0.00963
		20	0.08551	0.08112	0.07647	0.07152	0.06620	0.06037	0.05386	0.04626	0.03655
		30	0.19020	0.18145	0.17205	0.16187	0.15072	0.13832	0.12419	0.10739	0.08543
	40	10	0.06459	0.06121	0.05764	0.05384	0.04977	0.04534	0.04040	0.03466	0.02735
		20	0.17248	0.16439	0.15572	0.14636	0.13613	0.12480	0.11192	0.09666	0.07680
	50	10	0.11777	0.11191	0.10569	0.09902	0.09181	0.08389	0.07498	0.06453	0.05109

Table 5.4: Premium rates X^* that are the thresholds at which adverse selection will take place, for a variety of CI policies and initial wealth $W = £100,000$.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model III	20	10	0.00406	0.00217	0.00099	0.00037	0.00011	0.00002	0.00000	0.00000	0.00000
		20	0.01889	0.01022	0.00472	0.00178	0.00052	0.00011	0.00001	0.00000	0.00000
		30	0.06030	0.03363	0.01601	0.00625	0.00188	0.00040	0.00005	0.00000	0.00000
		40	0.13487	0.07933	0.03999	0.01657	0.00532	0.00120	0.00016	0.00001	0.00000
	30	10	0.01489	0.00803	0.00370	0.00139	0.00040	0.00008	0.00001	0.00000	0.00000
		20	0.05649	0.03141	0.01492	0.00580	0.00174	0.00037	0.00005	0.00000	0.00000
		30	0.13143	0.07712	0.03877	0.01602	0.00512	0.00115	0.00015	0.00001	0.00000
	40	10	0.04229	0.02328	0.01094	0.00421	0.00125	0.00026	0.00003	0.00000	0.00000
		20	0.11827	0.06876	0.03422	0.01399	0.00442	0.00098	0.00013	0.00001	0.00000
	50	10	0.07888	0.04458	0.02153	0.00852	0.00260	0.00056	0.00007	0.00000	0.00000
Model IV	20	10	0.00414	0.00240	0.00132	0.00069	0.00034	0.00017	0.00008	0.00004	0.00002
		20	0.01926	0.01129	0.00625	0.00330	0.00167	0.00082	0.00039	0.00019	0.00009
		30	0.06148	0.03714	0.02119	0.01152	0.00603	0.00306	0.00152	0.00074	0.00035
		40	0.13745	0.08743	0.05264	0.03028	0.01679	0.00905	0.00478	0.00248	0.00127
	30	10	0.01518	0.00888	0.00490	0.00257	0.00130	0.00064	0.00031	0.00014	0.00007
		20	0.05760	0.03470	0.01974	0.01071	0.00559	0.00283	0.00140	0.00068	0.00032
		30	0.13394	0.08500	0.05105	0.02929	0.01620	0.00871	0.00458	0.00237	0.00121
	40	10	0.04312	0.02572	0.01448	0.00777	0.00401	0.00201	0.00098	0.00047	0.00022
		20	0.12055	0.07582	0.04512	0.02563	0.01403	0.00746	0.00388	0.00199	0.00100
	50	10	0.08042	0.04922	0.02846	0.01570	0.00833	0.00429	0.00216	0.00107	0.00052

Table 5.5: Losses at which adverse selection occurs with $\sigma_R = 1.291$, *i.e.* the $(-3, 0)$ subpopulation no longer purchase at the rate offered by the insurer.

		Model			
Age	Term	I	II	III	IV
		£'s	£'s	£'s	£'s
20	10	45,600	25,100	3,100	3,100
	20	84,300	53,800	7,500	7,700
	30	100,000	61,600	9,100	9,400
	40	84,800	55,500	8,000	8,300
30	10	100,000	60,600	8,800	9,000
	20	100,000	63,800	9,500	9,900
	30	85,600	56,200	8,200	8,500
40	10	100,000	65,200	9,800	10,200
	20	85,300	55,800	8,100	8,300
50	10	80,300	50,600	7,000	7,200

5.3.3 Adverse Parameterisations of the Polygenic Model

We mentioned in Section 5.2.2 that Antoniou *et al.* (2002) estimated σ_R to be 1.291. This is the factor that allows us to transform the integer polygenotype value into a relative risk statistic (see Equation (5.2)). Essentially, σ_R measures the degree of risk dispersion arising from the polygene. We would expect that with wider dispersion of the intensities of BC (*i.e.* higher σ_R) there should be a greater chance of adverse selection. We will call the levels of σ_R which induce adverse selection *adverse parameterisations*.

Table 5.6 shows the levels of σ_R at which adverse selection would occur for all four utility models. Note that *both* premiums $X(-3, 0)$ and $\bar{X}$ are no longer fixed but vary depending on σ_R. The values in boldface correspond to adverse parameterisations which are lower than the Antoniou *et al.* estimate ($\hat{\sigma}_R$), hence correspond to a less severe polygene effect. As we would imagine, for higher losses individuals in subpopulation $(-3, 0)$ would be more motivated to insure and hence will 'adverse select' only at high values of σ_R where the polygene affects risk more severely.

The underlined figures in Table 5.6 are values at which the relative risk in the $P = +3$ subpopulation becomes exceptionally large and in turn the derivative of the reserve (Equation (1.2)) becomes infinite in terms of computation (*i.e.* greater than

the double-precision upper bound in C which is 1.797693×10^{308}). Therefore they are not adverse parameterisations, but bounds on our computations. We may assume that where this occurs individuals will *never* 'adverse select' since the BC intensities would have to be huge.

5.3.4 Adverse Selection by Multiple Subpopulations

Previously we considered the case where adverse selection is triggered when the lowest-risk subpopulation refuses to purchase insurance. However, given a population with such a broad range of risks, it may also be of interest to consider the prospect that more than one population will no longer purchase. Here we look at the situation where the subpopulations $(p, m) = (-3, 0)$ and $(-2, 0)$ adverse select, although we could easily extend this to more subpopulations. We propose two possibilities of modelling this:

(a) Adverse selection first occurs within the lowest risk subpopulation who as a result are removed from the pool of risks that the insurer must cover. The event of interest is now the point where the second lowest risk stops purchasing insurance at rate $\bar{X}_2 > \bar{X}$ which no longer includes the risks of the lowest risk subpopulation.

(b) If we assume that the two low risk populations act as one group in regards to their choice to purchase insurance, then the decision to refuse insurance will be based upon the average premium of these groups:

$$\frac{\omega(-3,0)X(-3,0) + \omega(-2,0)X(-2,0)}{\omega(-3,0) + \omega(-2,0)}. \tag{5.6}$$

Adverse selection occurs simultaneously in both genetic subpopulations when this pooled premium is below X^*.

Although the latter method is not strictly consistent with the economic model, if persons do know their individual risk, it is worthwhile considering in light of the comments we made in Section 2.3.4.

Table 5.6: Levels of σ_R at which adverse selection occurs, *i.e.* the (−3,0) subpopulation no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	20	10	**0.193**	**0.485**	**0.844**	**1.152**	1.393	1.592	1.770	1.942	2.130
		20	**0.050**	**0.131**	**0.232**	**0.363**	**0.530**	**0.733**	**0.957**	**1.189**	1.441
		30	**0.039**	**0.101**	**0.175**	**0.268**	**0.387**	**0.542**	**0.738**	**0.974**	**1.253**
		40	**0.048**	**0.119**	**0.207**	**0.319**	**0.467**	**0.658**	**0.889**	**1.151**	1.467
	30	10	**0.038**	**0.103**	**0.183**	**0.284**	**0.412**	**0.576**	**0.778**	**1.008**	**1.270**
		20	**0.036**	**0.094**	**0.163**	**0.249**	**0.359**	**0.501**	**0.686**	**0.914**	**1.192**
		30	**0.047**	**0.116**	**0.201**	**0.311**	**0.455**	**0.640**	**0.868**	**1.127**	1.441
	40	10	**0.036**	**0.092**	**0.159**	**0.242**	**0.347**	**0.483**	**0.659**	**0.880**	**1.153**
		20	**0.049**	**0.119**	**0.207**	**0.319**	**0.466**	**0.655**	**0.883**	**1.140**	1.445
	50	10	**0.060**	**0.146**	**0.256**	**0.399**	**0.583**	**0.803**	**1.040**	**1.286**	1.565
Model II	20	10	**0.462**	**1.073**	1.465	1.734	1.949	2.138	2.320	2.515	2.761
		20	**0.125**	**0.322**	**0.594**	**0.907**	**1.194**	1.447	1.685	1.934	2.259
		30	**0.096**	**0.238**	**0.432**	**0.685**	**0.970**	**1.248**	1.523	1.832	2.295
		40	**0.113**	**0.280**	**0.516**	**0.815**	**1.126**	1.432	1.767	2.210	3.136
	30	10	**0.098**	**0.253**	**0.463**	**0.732**	**1.014**	**1.277**	1.524	1.777	2.089
		20	**0.089**	**0.222**	**0.400**	**0.636**	**0.911**	**1.188**	1.463	1.766	2.209
		30	**0.110**	**0.273**	**0.502**	**0.795**	**1.104**	1.408	1.738	2.171	3.044
	40	10	**0.087**	**0.216**	**0.387**	**0.613**	**0.881**	**1.153**	1.421	1.708	2.101
		20	**0.113**	**0.281**	**0.516**	**0.813**	**1.120**	1.418	1.736	2.149	2.915
	50	10	**0.139**	**0.351**	**0.647**	**0.976**	**1.276**	1.551	1.837	2.196	2.767
Model III	20	10	2.263	2.930	3.705	4.867	7.529	7.529	7.529	7.529	7.529
		20	1.605	2.475	3.532	5.123	5.123	5.123	5.123	5.123	5.123
		30	1.412	2.536	4.606	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.587	3.377	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.444	2.292	3.245	5.401	5.401	5.401	5.401	5.401	5.401
		20	1.355	2.449	4.421	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.563	3.301	4.738	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	1.322	2.337	3.979	5.249	5.249	5.249	5.249	5.249	5.249
		20	1.574	3.204	4.898	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.708	3.039	5.176	5.176	5.176	5.176	5.176	5.176	5.176
Model IV	20	10	2.234	2.840	3.381	4.155	4.915	7.458	7.529	7.529	7.529
		20	1.566	2.346	3.127	4.244	5.123	5.123	5.123	5.123	5.123
		30	1.366	2.345	3.673	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.529	2.920	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.404	2.177	2.878	3.699	5.401	5.401	5.401	5.401	5.401
		20	1.309	2.266	3.452	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.506	2.855	4.733	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	**1.277**	2.175	3.109	4.810	5.249	5.249	5.249	5.249	5.249
		20	1.519	2.792	4.637	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.659	2.768	4.548	5.176	5.176	5.176	5.176	5.176	5.176

The threshold premiums calculated using the first method outlined above are given in Table 5.7 and Table 5.8. Adverse selection by the second subpopulation occurs when $X(-2,0) < X^*$ (the values of $X(-2,0)$ are given in Table 5.2). The threshold premiums for the second method are identical to those in Table 5.3 and Table 5.4, but the decision to insure is based on the result of Equation (5.6) which can be calculated from the figures in Table 5.2. Table 5.9 shows the adverse parameterisations calculated using the first method outlined above and Table 5.10 shows the adverse parameterisations calculated using the second method. In both tables the magnitudes of σ_R are greater than those in Table 5.6, since the polygene risks must be more severe to trigger adverse selection within a second subpopulation. The similarity of Table 5.9 and Table 5.10 suggests that with two adverse selecting subpopulations, the method by which individuals in those subpopulations evaluate their insurance-purchasing position is rather insignificant.

Table 5.7: Premium rates X^* that are the thresholds at which adverse selection by both the $P = -3$ and $P = -2$ polygenotype subpopulations will take place, for a variety of CI policies and initial wealth $W = £100,000$.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	20	10	0.00654	0.00636	0.00617	0.00596	0.00574	0.00549	0.00520	0.00486	0.00443
	20	20	0.03019	0.02937	0.02849	0.02755	0.02653	0.02539	0.02409	0.02254	0.02052
	20	30	0.09376	0.09137	0.08880	0.08601	0.08295	0.07952	0.07558	0.07085	0.06461
	20	40	0.19981	0.19527	0.19032	0.18488	0.17883	0.17197	0.16397	0.15421	0.14108
	30	10	0.02388	0.02323	0.02253	0.02178	0.02097	0.02006	0.01903	0.01781	0.01621
	30	20	0.08808	0.08582	0.08340	0.08076	0.07787	0.07464	0.07094	0.06649	0.06061
	30	30	0.19514	0.19068	0.18582	0.18049	0.17455	0.16783	0.16000	0.15045	0.13762
	40	10	0.06658	0.06484	0.06297	0.06095	0.05874	0.05627	0.05344	0.05006	0.04561
	40	20	0.17705	0.17292	0.16843	0.16351	0.15805	0.15188	0.14472	0.13600	0.12433
	50	10	0.12102	0.11802	0.11478	0.11126	0.10738	0.10302	0.09799	0.09193	0.08389
Model II	20	10	0.00637	0.00602	0.00565	0.00526	0.00485	0.00441	0.00391	0.00335	0.00263
	20	20	0.02943	0.02783	0.02616	0.02439	0.02251	0.02046	0.01820	0.01558	0.01227
	20	30	0.09156	0.08688	0.08193	0.07665	0.07097	0.06475	0.05778	0.04965	0.03924
	20	40	0.19565	0.18670	0.17708	0.16666	0.15523	0.14251	0.12800	0.11072	0.08811
	30	10	0.02327	0.02200	0.02067	0.01927	0.01778	0.01616	0.01436	0.01229	0.00968
	30	20	0.08599	0.08158	0.07691	0.07193	0.06657	0.06072	0.05417	0.04653	0.03677
	30	30	0.19105	0.18226	0.17283	0.16261	0.15142	0.13897	0.12478	0.10790	0.08584
	40	10	0.06497	0.06157	0.05798	0.05416	0.05007	0.04561	0.04064	0.03487	0.02752
	40	20	0.17325	0.16513	0.15643	0.14703	0.13676	0.12538	0.11245	0.09712	0.07717
	50	10	0.11825	0.11237	0.10613	0.09944	0.09220	0.08425	0.07530	0.06481	0.05131

Table 5.8: Premium rates X^* that are the thresholds at which adverse selection by both the $P = -3$ and $P = -2$ polygenotype subpopulations will take place, for a variety of CI policies and initial wealth $W = £100,000$.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model III	20	10	0.00407	0.00218	0.00100	0.00037	0.00011	0.00002	0.00000	0.00000	0.00000
	20	20	0.01898	0.01027	0.00474	0.00179	0.00052	0.00011	0.00001	0.00000	0.00000
	20	30	0.06064	0.03382	0.01611	0.00628	0.00189	0.00040	0.00005	0.00000	0.00000
	20	40	0.13551	0.07974	0.04022	0.01667	0.00535	0.00121	0.00016	0.00001	0.00000
	30	10	0.01497	0.00808	0.00372	0.00140	0.00041	0.00008	0.00001	0.00000	0.00000
	30	20	0.05682	0.03160	0.01501	0.00584	0.00175	0.00037	0.00005	0.00000	0.00000
	30	30	0.13206	0.07752	0.03900	0.01612	0.00516	0.00116	0.00015	0.00001	0.00000
	40	10	0.04255	0.02342	0.01101	0.00423	0.00126	0.00026	0.00003	0.00000	0.00000
	40	20	0.11884	0.06911	0.03442	0.01407	0.00445	0.00099	0.00013	0.00001	0.00000
	50	10	0.07922	0.04478	0.02164	0.00857	0.00262	0.00056	0.00007	0.00000	0.00000
Model IV	20	10	0.00415	0.00241	0.00132	0.00069	0.00035	0.00017	0.00008	0.00004	0.00002
	20	20	0.01936	0.01135	0.00628	0.00331	0.00168	0.00083	0.00040	0.00019	0.00009
	20	30	0.06183	0.03736	0.02132	0.01160	0.00607	0.00308	0.00153	0.00074	0.00036
	20	40	0.13809	0.08788	0.05294	0.03046	0.01690	0.00912	0.00481	0.00250	0.00128
	30	10	0.01527	0.00893	0.00493	0.00259	0.00131	0.00064	0.00031	0.00014	0.00007
	30	20	0.05793	0.03491	0.01987	0.01078	0.00562	0.00285	0.00141	0.00068	0.00033
	30	30	0.13458	0.08544	0.05135	0.02947	0.01631	0.00877	0.00462	0.00239	0.00122
	40	10	0.04338	0.02587	0.01457	0.00782	0.00404	0.00202	0.00099	0.00047	0.00022
	40	20	0.12113	0.07621	0.04537	0.02578	0.01412	0.00751	0.00391	0.00200	0.00101
	50	10	0.08076	0.04944	0.02860	0.01578	0.00838	0.00432	0.00217	0.00108	0.00052

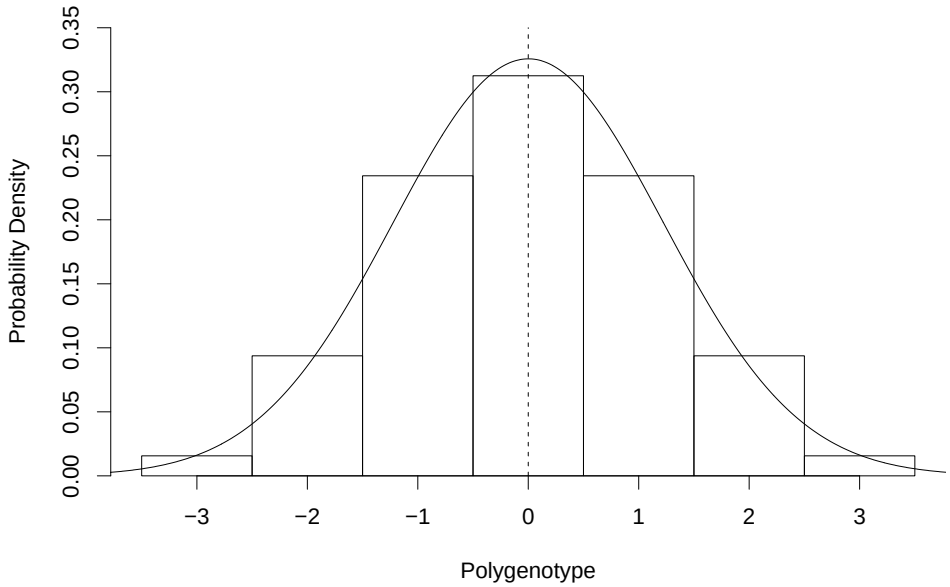


Figure 5.2: The binomial distribution with parameters (1/2, 6) (adjusted to have the mean at zero) overlaid with the Normal distribution with mean 0 and variance 3/2.

5.3.5 The Polygenotype as a Continuous Random Variable

In order to handle polygenic transmission (the passing down of genes) it is convenient to consider the polygenotype as a combination of individual genes (see Lange (1997)). However, polygenic theory (Strachan & Read, 2004) assumes that the range of risk derived from the polygene is continuous and usually that the distribution of the polygenotype is Normal. Since we do not need to consider inheritance of the polygene here, we no longer need to employ the binomial distribution as a discrete approximation of the Normally distributed polygenotype. We can assume that P is Normally distributed with parameters found from equating moments of the binomial distribution, hence $\mu = 0$ and $\sigma^2 = n\ q\ (1-q) = 2 \times 3 \times 1/2 \times 1/2 = 3/2$. The approximation that was made by using the binomial distribution to describe polygenotype frequency is shown in Figure 5.2 along with the Normal distribution that we will now use.

Now that we assume P is a continuous random variable the premium rates that depend on P are altered. The probability density function (p.d.f.) for $P \sim \mathcal{N}(0, 3/2)$ is $\omega_P^c(p)$, such that $\omega^c(p, m) = \omega_P^c(p)\omega_M(m)$ and the insurer's fair value premium is:

Table 5.9: Levels of σ_R at which adverse selection occurs within the $(-2, 0)$ subpopulation, *i.e.* the $(-3, 0)$ and $(-2, 0)$ subpopulations no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	20	10	**0.266**	**0.592**	**0.914**	**1.184**	1.405	1.595	1.768	1.938	2.124
		20	**0.072**	**0.184**	**0.315**	**0.467**	**0.639**	**0.824**	**1.019**	**1.225**	1.457
		30	**0.057**	**0.143**	**0.243**	**0.360**	**0.497**	**0.655**	**0.835**	**1.041**	**1.290**
		40	**0.069**	**0.168**	**0.284**	**0.421**	**0.581**	**0.764**	**0.968**	**1.199**	1.489
	30	10	**0.054**	**0.146**	**0.253**	**0.378**	**0.522**	**0.685**	**0.866**	**1.067**	1.300
		20	**0.052**	**0.134**	**0.228**	**0.338**	**0.466**	**0.616**	**0.789**	**0.990**	**1.236**
		30	**0.068**	**0.164**	**0.277**	**0.411**	**0.568**	**0.747**	**0.949**	**1.178**	1.465
	40	10	**0.052**	**0.131**	**0.222**	**0.328**	**0.452**	**0.597**	**0.765**	**0.961**	**1.200**
		20	**0.070**	**0.168**	**0.284**	**0.420**	**0.579**	**0.760**	**0.962**	**1.188**	1.468
	50	10	**0.086**	**0.205**	**0.344**	**0.507**	**0.690**	**0.888**	**1.094**	1.315	1.576
Model II	20	10	**0.568**	**1.113**	1.473	1.733	1.944	2.132	2.313	2.507	2.752
		20	**0.176**	**0.422**	**0.700**	**0.976**	**1.229**	1.463	1.689	1.932	2.251
		30	**0.136**	**0.323**	**0.544**	**0.788**	**1.037**	**1.285**	1.539	1.835	2.285
		40	**0.159**	**0.375**	**0.629**	**0.902**	**1.176**	1.455	1.772	2.199	3.086
	30	10	**0.139**	**0.340**	**0.575**	**0.826**	**1.072**	1.307	1.537	1.780	2.086
		20	**0.127**	**0.303**	**0.511**	**0.744**	**0.987**	**1.232**	1.484	1.772	2.201
		30	**0.156**	**0.367**	**0.616**	**0.885**	**1.157**	1.433	1.745	2.162	2.997
	40	10	**0.125**	**0.295**	**0.497**	**0.723**	**0.961**	**1.200**	1.444	1.716	2.097
		20	**0.160**	**0.375**	**0.629**	**0.900**	**1.170**	1.441	1.743	2.140	2.876
	50	10	**0.195**	**0.455**	**0.749**	**1.038**	1.306	1.563	1.837	2.186	2.745
Model III	20	10	2.256	2.921	3.687	4.845	7.529	7.529	7.529	7.529	7.529
		20	1.612	2.465	3.509	5.123	5.123	5.123	5.123	5.123	5.123
		30	1.435	2.521	4.551	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.600	3.324	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.461	2.285	3.227	5.401	5.401	5.401	5.401	5.401	5.401
		20	1.383	2.436	4.366	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.578	3.248	4.738	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	1.352	2.328	3.917	5.249	5.249	5.249	5.249	5.249	5.249
		20	1.587	3.153	4.898	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.712	3.006	5.176	5.176	5.176	5.176	5.176	5.176	5.176
Model IV	20	10	2.227	2.831	3.368	4.137	4.892	7.375	7.529	7.529	7.529
		20	1.575	2.337	3.109	4.178	5.123	5.123	5.123	5.123	5.123
		30	1.392	2.334	3.615	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.545	2.878	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.424	2.171	2.863	3.673	5.401	5.401	5.401	5.401	5.401
		20	1.340	2.257	3.400	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.524	2.816	4.658	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	1.312	2.168	3.082	4.756	5.249	5.249	5.249	5.249	5.249
		20	1.535	2.759	4.571	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.665	2.746	4.493	5.176	5.176	5.176	5.176	5.176	5.176

Table 5.10: Levels of σ_R at which adverse selection occurs when subpopulations $(-3, 0)$ and $(-2, 0)$ pool their premium, *i.e.* the (-3,0) and (-2,0) subpopulations no longer purchase at the rate offered by the insurer. Figures in bold correspond to parameterisations lower than in the fitted model of Antoniou *et al.* (2002). Figures underlined produce relative risk statistics that result in numerical overflows.

			Loss to Wealth Ratio								
	Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Model I	20	10	**0.259**	**0.587**	**0.916**	**1.190**	1.414	1.604	1.778	1.947	2.133
		20	**0.070**	**0.178**	**0.308**	**0.460**	**0.634**	**0.823**	**1.023**	**1.231**	1.465
		30	**0.055**	**0.138**	**0.236**	**0.352**	**0.489**	**0.649**	**0.834**	**1.044**	1.297
		40	**0.067**	**0.162**	**0.276**	**0.413**	**0.574**	**0.761**	**0.970**	**1.205**	1.500
	30	10	**0.052**	**0.141**	**0.246**	**0.370**	**0.514**	**0.680**	**0.866**	**1.070**	1.307
		20	**0.050**	**0.129**	**0.221**	**0.329**	**0.458**	**0.610**	**0.787**	**0.992**	**1.241**
		30	**0.065**	**0.158**	**0.270**	**0.403**	**0.561**	**0.744**	**0.951**	**1.184**	1.475
	40	10	**0.050**	**0.126**	**0.215**	**0.320**	**0.444**	**0.591**	**0.762**	**0.962**	**1.205**
		20	**0.067**	**0.163**	**0.276**	**0.412**	**0.573**	**0.758**	**0.964**	**1.194**	1.478
	50	10	**0.083**	**0.198**	**0.336**	**0.500**	**0.686**	**0.888**	**1.099**	1.323	1.587
Model II	20	10	**0.563**	**1.119**	1.482	1.742	1.953	2.141	2.322	2.516	2.761
		20	**0.170**	**0.414**	**0.696**	**0.978**	**1.236**	1.471	1.698	1.942	2.263
		30	**0.131**	**0.315**	**0.537**	**0.785**	**1.040**	1.291	1.549	1.848	2.304
		40	**0.154**	**0.367**	**0.624**	**0.903**	**1.183**	1.466	1.788	2.222	3.141
	30	10	**0.135**	**0.332**	**0.568**	**0.824**	**1.076**	1.314	1.545	1.789	2.096
		20	**0.123**	**0.295**	**0.503**	**0.741**	**0.989**	**1.238**	1.493	1.784	2.218
		30	**0.150**	**0.358**	**0.610**	**0.885**	**1.163**	1.443	1.760	2.183	3.049
	40	10	**0.120**	**0.287**	**0.489**	**0.719**	**0.962**	**1.205**	1.452	1.726	2.111
		20	**0.155**	**0.367**	**0.623**	**0.900**	**1.176**	1.451	1.757	2.161	2.921
	50	10	**0.189**	**0.448**	**0.746**	**1.042**	1.313	1.573	1.851	2.204	2.771
Model III	20	10	2.265	2.930	3.705	4.867	7.529	7.529	7.529	7.529	7.529
		20	1.621	2.479	3.533	5.123	5.123	5.123	5.123	5.123	5.123
		30	1.444	2.542	4.606	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.613	3.380	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.469	2.297	3.246	5.401	5.401	5.401	5.401	5.401	5.401
		20	1.391	2.456	4.421	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.590	3.305	4.738	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	1.359	2.345	3.980	5.249	5.249	5.249	5.249	5.249	5.249
		20	1.600	3.208	4.898	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.724	3.042	5.176	5.176	5.176	5.176	5.176	5.176	5.176
Model IV	20	10	2.236	2.840	3.381	4.155	4.915	7.458	7.529	7.529	7.529
		20	1.584	2.350	3.128	4.245	5.123	5.123	5.123	5.123	5.123
		30	1.400	2.353	3.675	4.811	4.811	4.811	4.811	4.811	4.811
		40	1.557	2.926	4.626	4.626	4.626	4.626	4.626	4.626	4.626
	30	10	1.431	2.182	2.880	3.699	5.401	5.401	5.401	5.401	5.401
		20	1.348	2.274	3.455	4.971	4.971	4.971	4.971	4.971	4.971
		30	1.536	2.861	4.733	4.738	4.738	4.738	4.738	4.738	4.738
	40	10	1.318	2.184	3.112	4.811	5.249	5.249	5.249	5.249	5.249
		20	1.547	2.798	4.637	4.898	4.898	4.898	4.898	4.898	4.898
	50	10	1.677	2.772	4.548	5.176	5.176	5.176	5.176	5.176	5.176

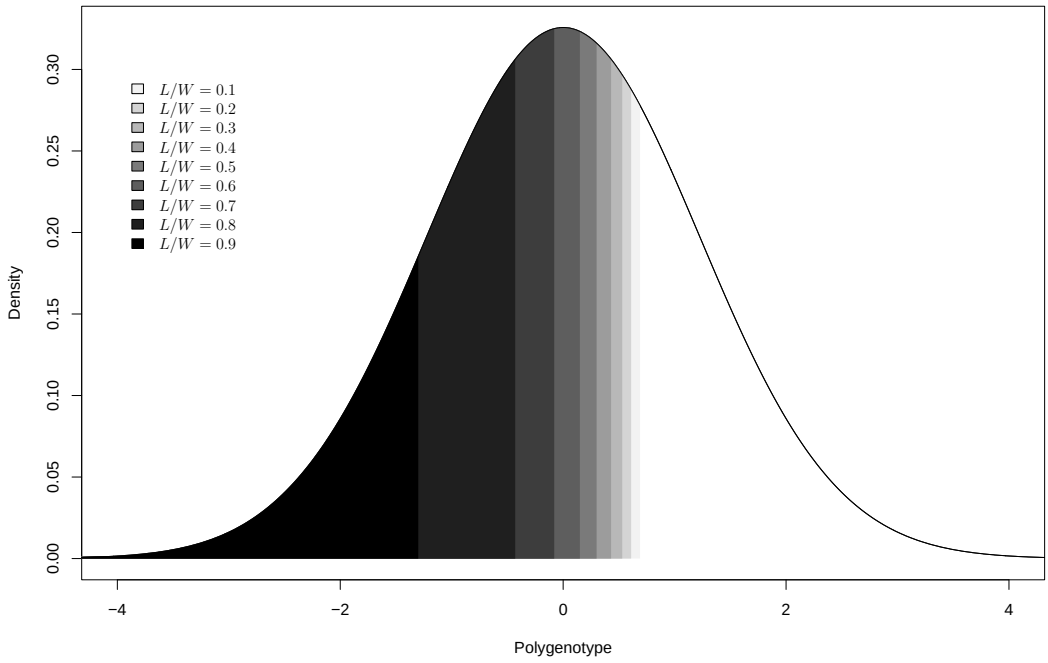


Figure 5.3: The Normal polygenotype distribution in the BRCA0 subpopulation. The proportions who adverse select on a 10-year term-assurance beginning at age 40 under the assumption of Model I utility are shaded in a series of overlapping greys corresponding to the loss to wealth ratio.

$$\bar{X}^c = \sum_m \int_{-\infty}^{\infty} X(p,m)\omega^c(p,m)dp. \tag{5.7}$$

We can use this new format to find the value of p^* in the BRCA0 subpopulation where the adverse selection threshold exists. Also, by integrating $\omega(p,0)$ from $-\infty$ to the threshold value of p^*, we can find the proportion of individuals who decline to purchase CI insurance. Figure 5.3 shows the answer graphically for a 10-year term-assurance with an entry age of 40 and Model I utility.

Our results for all policies are given in Table 5.11 and Table 5.12. Clearly the effect of adverse selection is at its greatest when possible losses are relatively low, and we can see that even risky (positive) polygenotype carriers 'adverse select' at sufficiently low levels of loss. Note that risk aversion in Model III is too high to lead to adverse selection at any level of loss and any policy type. However, in Model IV there are two instances where adverse selection may arise: a policy with entry age 30 and term 20

years has $p^* = -3.95$, representing 0.1% of the market, and a policy with entry age 40 and term 10 years has $p^* = -2.60$, representing 1.7%.

Table 5.11: The polygenotype p^* at which adverse selection occurs for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model I utility. The figures in parentheses represent the proportion of the market who will not purchase insurance.

		Loss to Wealth Ratio								
Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	10	0.56 (67.5%)	0.28 (58.9%)	-0.14 (45.3%)	-1.00 (20.7%)	$-\infty$ (0.0%)	$-\infty$ (0.0%)	$-\infty$ (0.0%)	$-\infty$ (0.0%)	$-\infty$ (0.0%)
	20	0.7 (71.4%)	0.60 (68.6%)	0.49 (65.4%)	0.35 (61.1%)	0.17 (55.4%)	-0.08 (47.3%)	-0.48 (34.7%)	-1.39 (12.8%)	$-\infty$ (0.0%)
	30	0.67 (70.6%)	0.59 (68.3%)	0.50 (65.7%)	0.39 (62.3%)	0.25 (57.9%)	0.07 (52.2%)	-0.19 (43.7%)	-0.64 (30.0%)	-2.16 (3.9%)
	40	0.62 (69.2%)	0.53 (66.6%)	0.42 (63.3%)	0.29 (59.2%)	0.12 (53.8%)	-0.11 (46.3%)	-0.49 (34.4%)	-1.33 (13.8%)	$-\infty$ (0.0%)
30	10	0.71 (71.7%)	0.63 (69.5%)	0.54 (66.9%)	0.42 (63.3%)	0.28 (58.9%)	0.10 (53.1%)	-0.17 (44.4%)	-0.64 (30.0%)	-2.38 (2.6%)
	20	0.67 (70.6%)	0.60 (68.6%)	0.51 (66.0%)	0.41 (63.0%)	0.28 (58.9%)	0.11 (53.5%)	-0.12 (46.0%)	-0.51 (33.8%)	-1.56 (10.1%)
	30	0.62 (69.2%)	0.53 (66.6%)	0.43 (63.6%)	0.30 (59.5%)	0.14 (54.4%)	-0.09 (47.0%)	-0.45 (35.6%)	-1.22 (15.9%)	$-\infty$ (0.0%)
40	10	0.69 (71.2%)	0.61 (68.9%)	0.53 (66.6%)	0.43 (63.6%)	0.30 (59.5%)	0.15 (54.7%)	-0.08 (47.3%)	-0.43 (36.2%)	-1.30 (14.4%)
	20	0.63 (69.5%)	0.54 (66.9%)	0.43 (63.6%)	0.30 (59.5%)	0.14 (54.4%)	-0.10 (46.6%)	-0.46 (35.3%)	-1.26 (15.1%)	$-\infty$ (0.0%)
50	10	0.65 (70.1%)	0.54 (66.9%)	0.42 (63.3%)	0.27 (58.6%)	0.06 (51.8%)	-0.24 (42.1%)	-0.76 (26.7%)	-2.79 (1.1%)	$-\infty$ (0.0%)

Table 5.12: The polygenotype p^* at which adverse selection occurs for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model II utility. The figures in parentheses represent the proportion of the market who will not purchase insurance.

		Loss to Wealth Ratio								
Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	10	0.31	-0.66	-∞	-∞	-∞	-∞	-∞	-∞	-∞
		(59.8%)	(29.4%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
	20	0.61	0.39	0.10	-0.37	-1.43	-∞	-∞	-∞	-∞
		(68.9%)	(62.3%)	(53.1%)	(38.0%)	(12.1%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
	30	0.60	0.43	0.20	-0.11	-0.63	-2.09	-∞	-∞	-∞
		(68.6%)	(63.6%)	(56.4%)	(46.3%)	(30.3%)	(4.4%)	(0.0%)	(0.0%)	(0.0%)
	40	0.54	0.34	0.07	-0.35	-1.20	-∞	-∞	-∞	-∞
		(66.9%)	(60.8%)	(52.2%)	(38.7%)	(16.3%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
30	10	0.64	0.46	0.23	-0.10	-0.66	-2.55	-∞	-∞	-∞
		(69.8%)	(64.5%)	(57.3%)	(46.6%)	(29.4%)	(1.9%)	(0.0%)	(0.0%)	(0.0%)
	20	0.61	0.44	0.23	-0.05	-0.50	-1.53	-∞	-∞	-∞
		(68.9%)	(63.9%)	(57.3%)	(48.3%)	(34.1%)	(10.6%)	(0.0%)	(0.0%)	(0.0%)
	30	0.54	0.35	0.08	-0.32	-1.11	-∞	-∞	-∞	-∞
		(66.9%)	(61.1%)	(52.5%)	(39.6%)	(18.2%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
40	10	0.62	0.46	0.26	-0.01	-0.43	-1.30	-∞	-∞	-∞
		(69.2%)	(64.5%)	(58.3%)	(49.6%)	(36.2%)	(14.4%)	(0.0%)	(0.0%)	(0.0%)
	20	0.55	0.35	0.08	-0.33	-1.15	-∞	-∞	-∞	-∞
		(67.2%)	(61.1%)	(52.5%)	(39.3%)	(17.3%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
50	10	0.55	0.32	-0.02	-0.58	-2.52	-∞	-∞	-∞	-∞
		(67.2%)	(60.2%)	(49.2%)	(31.7%)	(2.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)

An important assumption that we made here is that the lower limit of integration in Equation (5.7) is $-\infty$ and hence does not adapt as those with low polygenotypes leave the market. This implies that the insurance company is unaware of how the risk-pool is affected by the advent of adverse selection and does not adapt their premiums accordingly. This seems plausible given that we are assuming the insurer does not receive any genetic information about the population. However, it is possible to assume that the insurer is eventually able to update their assessment of the risks (while still unable to distinguish between them) in the insurance-purchasing population and therefore offer a premium that is representative of this group.

We can adapt our method to consider an insurance market which adapts dynamically to the risk-pool. This means that now the fair value premium (per unit of loss) that the insurer will set is:

$$\bar{X}^c = \frac{\int_{p^*}^{\infty} X(p,0)\omega^c(p,0)dp + \int_{-\infty}^{\infty} X(p,1)\omega^c(p,1)dp + \int_{-\infty}^{\infty} X(p,2)\omega^c(p,2)dp}{1 - \int_{-\infty}^{p^*} \omega^c(p,0)dp}. \quad (5.8)$$

We shall call this the 'dynamic setting', which serves as a conceivable alternative to the previous 'static setting'.

As before, we find the values of the polygenotype p^*, where those with polygenotype less than p^* refuse to purchase insurance, but now assuming the insurer can adapt their premiums in accordance with the risks remaining in the insurance pool after adverse selection. These thresholds are given in Table 5.13 and Table 5.14 and are markedly higher than those found previously. Note that there are instances where nearly the entire BRCA0 subpopulation adverse selects. This could easily 'spill over' into the BRCA1 and BRCA2 subpopulations but for purposes of demonstration it suffices to show that the vast majority of the population adverse selects. Such behaviour would spell disaster for the insurer's CI business. We find that adverse selection does not occur commonly under utility models III and IV. In Model III a policy with entry age 30 and term 10 has $p^* = -3.74$ representing 0.1% of the population. In Model IV a policy with entry age 30 and term 20 has $p^* = -2.86$ representing 1% of the

population and a policy with entry age 30 and term 10 has $p^* = 2.74$ representing 98.5% of the population. Therefore we find that adverse selection can potentially be a serious problem for some insurance products.

Table 5.13: The polygenotype p^* at which adverse selection occurs under the dynamic insurer pricing method for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model I utility. The figures in parentheses represent the proportion of the market who will not purchase insurance.

		Loss to Wealth Ratio								
Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	10	3.25 (99.4%)	3.21 (99.3%)	3.15 (99.3%)	3.09 (99.2%)	-∞ (0.0%)	-∞ (0.0%)	-∞ (0.0%)	-∞ (0.0%)	-∞ (0.0%)
	20	3.63 (99.6%)	3.60 (99.6%)	3.57 (99.6%)	3.53 (99.6%)	3.49 (99.5%)	3.43 (99.5%)	3.34 (99.4%)	3.22 (99.3%)	-∞ (0.0%)
	30	3.40 (99.5%)	3.38 (99.5%)	3.35 (99.5%)	3.32 (99.4%)	3.27 (99.4%)	3.22 (99.3%)	3.14 (99.2%)	3.00 (99.0%)	2.68 (98.3%)
	40	3.25 (99.4%)	3.23 (99.3%)	3.20 (99.3%)	3.17 (99.3%)	3.13 (99.2%)	3.08 (99.2%)	3.00 (99.0%)	2.86 (98.8%)	-∞ (0.0%)
30	10	3.68 (99.6%)	3.65 (99.6%)	3.62 (99.6%)	3.58 (99.6%)	3.54 (99.6%)	3.48 (99.5%)	3.40 (99.5%)	3.29 (99.4%)	3.06 (99.1%)
	20	3.41 (99.5%)	3.39 (99.5%)	3.36 (99.5%)	3.33 (99.4%)	3.28 (99.4%)	3.23 (99.3%)	3.15 (99.3%)	3.02 (99.1%)	2.72 (98.4%)
	30	3.25 (99.4%)	3.23 (99.3%)	3.21 (99.3%)	3.18 (99.3%)	3.14 (99.2%)	3.08 (99.2%)	3.01 (99.1%)	2.87 (98.8%)	-∞ (0.0%)
40	10	3.42 (99.5%)	3.40 (99.5%)	3.37 (99.5%)	3.33 (99.4%)	3.29 (99.4%)	3.23 (99.3%)	3.15 (99.3%)	3.02 (99.1%)	2.73 (98.5%)
	20	3.26 (99.4%)	3.24 (99.4%)	3.21 (99.3%)	3.18 (99.3%)	3.14 (99.2%)	3.09 (99.2%)	3.01 (99.1%)	2.86 (98.8%)	-∞ (0.0%)
50	10	3.32 (99.4%)	3.30 (99.4%)	3.27 (99.4%)	3.23 (99.3%)	3.18 (99.3%)	3.12 (99.2%)	3.03 (99.1%)	2.87 (98.8%)	-∞ (0.0%)

Table 5.14: The polygenotype p^* at which adverse selection occurs under the dynamic insurer pricing method for a variety of policy entry ages and terms, with $\sigma_R = 1.291$, $W = £100,000$ and Model II utility. The figures in parentheses represent the proportion of the market who will not purchase insurance.

		Loss to Wealth Ratio								
Age	Term	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	10	3.21	3.11	-∞	-∞	-∞	-∞	-∞	-∞	-∞
		(99.3%)	(99.2%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
	20	3.61	3.55	3.47	3.37	3.23	-∞	-∞	-∞	-∞
		(99.6%)	(99.6%)	(99.5%)	(99.5%)	(99.3%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
	30	3.38	3.33	3.27	3.18	3.05	2.81	-∞	-∞	-∞
		(99.5%)	(99.4%)	(99.4%)	(99.3%)	(99.1%)	(98.7%)	(0.0%)	(0.0%)	(0.0%)
	40	3.23	3.18	3.13	3.05	2.93	-∞	-∞	-∞	-∞
		(99.3%)	(99.3%)	(99.2%)	(99.1%)	(98.9%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
30	10	3.66	3.60	3.52	3.43	3.30	3.08	-∞	-∞	-∞
		(99.6%)	(99.6%)	(99.6%)	(99.5%)	(99.4%)	(99.2%)	(0.0%)	(0.0%)	(0.0%)
	20	3.39	3.34	3.28	3.19	3.06	2.83	-∞	-∞	-∞
		(99.5%)	(99.4%)	(99.4%)	(99.3%)	(99.1%)	(98.7%)	(0.0%)	(0.0%)	(0.0%)
	30	3.23	3.19	3.13	3.06	2.94	-∞	-∞	-∞	-∞
		(99.3%)	(99.3%)	(99.2%)	(99.1%)	(98.9%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
40	10	3.40	3.35	3.28	3.19	3.05	2.81	-∞	-∞	-∞
		(99.5%)	(99.5%)	(99.4%)	(99.3%)	(99.1%)	(98.7%)	(0.0%)	(0.0%)	(0.0%)
	20	3.24	3.20	3.14	3.06	2.94	-∞	-∞	-∞	-∞
		(99.4%)	(99.3%)	(99.2%)	(99.1%)	(98.9%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
50	10	3.30	3.24	3.17	3.08	2.92	-∞	-∞	-∞	-∞
		(99.4%)	(99.4%)	(99.3%)	(99.2%)	(98.9%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)

5.4 Conclusions

We calculated the parameterisations of the polygene model that would lead to adverse selection. Models III and IV are calibrated in a manner more representative of Italian risk preferences than models I and II and we found that under models III and IV the polygene would have to confer a very wide range of risk in order to effect adverse selection.

When we considered the polygenotype as a continuous random variable we could find the value of the polygenotype which divided those who would buy insurance from those who would not. For utility models I and II, even with the risk of large losses, large proportions of the population would be prepared to not purchase cover. When we made the insurers' premium calculation dynamic we saw these proportions increase sharply. For models III and IV adverse selection only took place in the presence of the highest risks of BC onset. These levels of adverse selection were uncommon and considerably low in the static-pricing setting, however adverse selection could become severe in the dynamic setting for one of the policies we considered (namely that with entry age 40 and term 10) under the utility of Model IV.

It is unlikely that adverse selection will arise except when individuals perceive their potential losses as low and have relatively low risk aversion. We found that under models III and IV testing for the polygene would be unlikely to deter low risk individuals from insurance.

Chapter 6

Longevity Genes

6.1 Pension Annuities and Genetics

6.1.1 Genes for Longevity

Over five decades, from 1953 to 2003, the number of people in the UK aged 50 and over increased by 45%, to 20 million. They are projected to number 27.2 million in 25 years (Office for National Statistics) and among them, a higher proportion will be at the oldest ages. This is splendid progress, but it will place strain on the financing of retirement income.

If genes were identified, that clearly identified higher mortality risk at older ages (we will call these 'frailty genes') then their carriers might (by what has been called 'actuarial fairness') be entitled to a higher rate of annuity. However, carriers of genes that confer longer life (we will call these 'longevity genes') would face more expensive annuities. Studies of identical and non-identical twins estimate the heritability of longevity (the variation in genes as a proportion of the variation in life expectancy) to be about 25%, enough to warrant further study (Herskind *et al.*, 1996; McGue *et al.*, 1993). Table 6.1 lists some of the genes that have been linked to longevity. Most of them are related to Alzheimer's disease (AD) and heart disorders.

In Table 6.1, mtDNA refers to mitochondrial DNA; the DNA that is found outside of the cell nucleus. We discuss mitochondrial DNA in greater detail in Section 1.1.2.

Table 6.1: Genes, and their possible related disorders, that have been repeatedly studied for associations with longevity and shown significant correlations (De Benedictis *et al.*, 2001).

Gene	Disease
ApoE	Alzheimer's disease, Cardiovascular disease
ApoB	Coronary artery disease
ApoA-IV	Alzheimer's disease
ACE	Myocardial infarction, Cerebral infarction, Alzheimer's disease, Essential hypertension
CYP2D6	Parkinson disease
HLA1 & HLA2	Immune Disorders
P53	Cancer
Factors V, VII	Myocardial infarction
Fibrinogen	Coronary artery disease
Prothrombin	Myocardial infarction
MTHFR	Cardiovascular disease, Cancer
mtDNA	Coronary artery disease, Diabetes, Parkinson disease, Alzheimer's disease
PARP	Unknown

Some work has been completed that investigates how genetic testing impacts insurance products which cover the risks of elderly populations. Macdonald & Pritchard (2000) studied the APOE gene in relation to Alzhiemer's disease and how this may affect the long-term care insurance market.

6.1.2 'Disease Genes' and Longevity

Obvious candidates as frailty/longevity genes are those known to be important in developing disease. In some insurance contexts, this intuition could be misleading.

Macdonald, Waters & Wekwete (2005a, 2005b) showed that genes that confer high risk of developing major risk factors of heart disease (for example, hypercholesterolemia and hypertension) need not, by themselves, be good predictors of the disease itself. However, considering longevity instead of disease onset, the opposite is true. Genes that control risk factors, that in turn influence common diseases, are the important genetic markers for longevity. This is discussed in De Benedictis *et al.*, (2001), a review of longevity genetics, which concludes that the important longevity genes are not those that influence mortality solely through disease. The Apolipoprotein genes

are a good example; they play important rôles in the regulation of cholesterol and therefore can be instrumental in longevity.

6.1.3 Tan *et al.* (2001)

The bugbear of actuarial studies in genetics is that actuarial questions require age-related rates of disease onset, while many medical questions can be answered by simpler statistics. Thus Macdonald & Pritchard (2000) trawled the large literature on AD (up to about 1998), and found just one study that reported age-related risks in enough detail. In this light, the study by Tan *et al.* (2001) described below is particularly useful for two reasons: (a) it includes a wide selection of genes with varied effects on longevity; and (b) it reports its results in the form of relative risks, from which rates of onset can be extracted quite easily.

Tan *et al.* (2001) fitted a Cox proportional hazards model to estimate the influence of candidate frailty/longevity genes on lifespan. The study population comprised 961 Italians of whom 212 (22%) were centenarians. Many of the genes listed in Table 6.1 were missing from this study; some were not studied at all, while others that did not show statistically significant relative risks were dropped. On the other hand, some genes that did not attain significance on their own were kept because they showed significance in interaction with environment or gender. The genetic influence on longevity is almost certainly polygenic (arising from combinations of many genes, each of small influence) and clearly has a large environmental component. Hunting gene×gene interactions will perhaps be the most fruitful approach in the long term, for which the Cox model is very suitable. Tan *et al.* (2001) did not pursue this (just mentioning that such discoveries would be "interesting and necessary").

However, the sample sizes in Tan *et al.* (2001) vary (Table 6.2), and some are quite small, so we must consider the sampling errors of any premium rate based on their model. Surprisingly, given that medical studies have been the bedrock of medical underwriting for many years, this question is hardly ever asked. No underwriting manual, to our knowledge, shows confidence intervals for premium rates. Part of the reason is that without the data, it is difficult to estimate the sampling properties

Table 6.2: List of genes studied in Tan *et al.* (2001) labelled $g = 1, 2, \ldots, 12$; and the KLOTHO genotypes studied in Arking *et al.* (2005), labelled $g = 13, 14$.

g	Gene	Sample Size
1	Apob35	787
2	Apob39	787
3	THO7	555
4	THO8	555
5	THO10	555
6	SOD2-T	354
7	INS-	438
8	INS+	438
9	mtDNAhapl-J	547
10	mtDNAhapl-U	547
11	mtDNAstr-136	393
12	mtDNAstr-138	393
13	KLOTHO FF	216
14	KLOTHO VV	216

of any quantity. Lu, Macdonald & Waters (2007) did so in the case of summary data available from some non-parametric studies of polycystic kidney disease, but the opportunity was unusual. We will show that, thanks to the structure imposed by the Cox model, Tan *et al.* (2001) give sufficient information to approximate sampling distributions of premium rates.

The genes studied by Tan *et al.* (2001) were chosen by reviewing the longevity genetics literature back to as far as 1982. Most of these papers found associations by observing the population frequency of a gene at different ages. Increases (decreases) in the gene frequency at older ages would identify a longevity (frailty) gene, respectively. Tan *et al.* (2001) began with over 70 gene variants, but these were reduced to the first 12 shown in Table 6.2 after genes leading to no significant results were eliminated.

Like other genetic longevity studies, Tan *et al.* (2001) observed gene frequencies in an elderly group (cases) and in a younger group (controls). Because this was a cross-sectional study the Cox model could not be directly applied, however the Cox proportional hazards assumption was used to constrain the likelihood and increase power. A maximum likelihood approach was used to estimate the proportions in each of several subpopulation (defined by the binary covariates genotype, gender and

north/south region) and the relative risks in each of these subpopulations. Although the proportion p_i in the ith subpopulation was observed at the time the cohort was observed, and did not need to be estimated, it was also necessary to estimate the proportions $p_i(x)$ at each age x.

6.1.4 Arking *et al.* (2005)

Arking *et al.* (2005) studied the KLOTHO genotype. They did not consider simply the presence or absence of a single gene variant, but the possible combinations of two variants of the same gene (called 'alleles'). Denoting one allele F and the other V, we have three possible genotypes: FF, FV, and VV. Arking *et al.* studied the longevity of FF and VV carriers relative to FV carriers (FF carriers and VV carriers are denoted $g = 13, 14$, respectively, in Table 6.2 and throughout this paper. Arking *et al.* fitted a Cox model over approximately four years of follow-up.

6.2 Parameter Uncertainty in the Cox Model

6.2.1 The Cox Model

The ubiquitous Cox model is a semi-parametric multiplicative hazard regression model. Let t be a suitable timescale (such as age). Individual i ($i = 1, 2, \ldots, n$) has force of mortality $\lambda^i(t)$ of the form:

$$\lambda^i(t; \boldsymbol{Z}_i) = \lambda^0(t) \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \tag{6.1}$$

where: $\boldsymbol{Z}_i$ is a p-dimensional vector of covariates (risk factors) for individual i; $\lambda^0(t)$ is the baseline force of mortality; and $\boldsymbol{\beta}$ is the p-dimensional vector of regression coefficients. Andersen *et al.* (1993) is a definitive reference on hazard regression models.

6.2.2 Parameter Uncertainty in the Cox Model

In the Cox model the fitted parameter is $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$. Sometimes the baseline hazard $\lambda^0(t)$ is estimated as well. In either case, the fitting process usually yields standard errors for the parameter estimates, thus parameter uncertainty can be quantified. Other, important, forms of uncertainty, such as stochastic uncertainty and model uncertainty (see Cairns (2000)), lie outside the model.

Tan *et al.* (2001) published estimates $\widehat{RR}_g$ of the relative risk RR_g in respect of each gene labelled $g = 1, 2, \ldots, 12$ in Table 6.2. The relative risk is just the multiplier of the baseline hazard, $RR_g = \exp(\boldsymbol{\beta}^\top \boldsymbol{z}_i)$. They modelled each gene separately, not all twelve simultaneously, so in each case the covariate vector for individual i included a single binary component z_{ig} indicating genotype. Estimation was by maximum likelihood, using a two-step procedure adapted to their non-standard sampling scheme (the relevant information lay in the proportion of individuals alive at the time of study). The sampling (parameter) uncertainty in the estimates $\widehat{RR}_g$ was supplied, in the form of sample standard deviations. We can use these to estimate the sampling distributions of premium rates based on the fitted hazard rates.

6.2.3 A Remark on the Baseline Hazards

Our model differs from Tan *et al.* (2001) in the derivation of the baseline hazard function. They used Italian population mortality statistics from 1994 (male and female) as initial estimates of the baseline hazard. They then used a two-step algorithm.

(a) Step one was to condition on this baseline to obtain conditional MLEs of subpopulation proportions at all ages x, and relative risks.

(b) Step two was to calculate a new baseline hazard, as a weighted average hazard rate, the weights being the proportions in each subpopulation.

These steps were repeated until convergence was achieved. The authors preferred this approach to a model stratified by sex, because they wished to model gene $\times$ sex interactions, which needed a single baseline hazard. They did not publish the baseline

hazard they used, so as a proxy, we have used Italian male and female population mortality in 1994 life tables available online to obtain a hazard rate as follows:

$$\lambda^0(t) = \frac{l_f(t)\ \lambda_f(t) + l_m(t)\ \lambda_m(t)}{l_f(t) + l_m(t)} \tag{6.2}$$

where $l_f(t)$ and $l_m(t)$ are the standard life table functions for the expected numbers alive at time t, for females and males respectively. Note that $l_f(t) = l_m(t)$ at time $t = 0$, but $l_f(t) > l_m(t)$ for $t > 0$.

Arking *et al.* (2005) studied a small cohort of 216 Ashkenazi Jews in the U.S.A., so in this case we approximated the baseline hazard using USA life table data from 2000.

6.2.4 Sampling Distributions of Relative Risks and Premiums

Given a relative risk estimate $\widehat{RR}_g$ and the baseline hazard $\lambda^0(t)$, it is simple to calculate the single premium for a whole-life annuity of 1 per year, payable continuously; denote this $\widehat{P}_g$. The notation emphasises that if we knew the true relative risk RR_g we could compute the true premium rate denoted P_g, but in practice we only obtain the point estimate $\widehat{P}_g$. We could express the premium rate as a function of the relative risk: $P_g = f(RR_g)$ and through this relationship $\widehat{P}_g$ inherits a sampling distribution from that of $\widehat{RR}_g$. This is our real target of study. The simplest way to find it, given that $f()$ is a somewhat complicated function, is by simulating from the sampling distribution of $\widehat{RR}_g$.

Tan *et al.* (2001) estimated the standard errors of the $\widehat{RR}_g$. Assuming the estimate $\hat{\boldsymbol{\beta}}$ to be multivariate Normal (justified asymptotically), $\widehat{RR}_g$ has a log-normal distribution, with parameters μ_g and σ_g. Hence, given the estimate $\widehat{RR}_g$ and its estimated standard deviation $S[\widehat{RR}_g]$ we can find μ_g and σ_g by equating first and second moments:

$$\mu_g = \log\left(\frac{\widehat{RR}_g^2}{\sqrt{\widehat{RR}_g^2 + S[\widehat{RR}_g]^2}}\right) \tag{6.3}$$

$$\sigma_g^2 = \log\left(\left(\frac{S[\widehat{RR}_g]}{\widehat{RR}_g}\right)^2 + 1\right). \tag{6.4}$$

The sampling density functions of $\widehat{RR}_g$ for each gene g are shown in Figure 6.1.

We also tried Gamma sampling distributions for the relative risks. They gave very similar results, and Gamma distributions lack the multiplicative property of log-normal distributions (which will be useful in Section 6.2.6) so we did not pursue them further. The sampling densities for females' relative risk when assuming a Gamma distribution are shown in Figure 6.2 for comparison. Differences between the two distributions are only visible at relative risks of low magnitude. For example, compare the density of the risk reducing mtDNAstr-138 allele in Figure 6.1 with it's counterpart in Figure 6.2.

6.2.5 Premiums for Females

We simulated 10,000 samples from the sampling distributions of each $\widehat{RR}_g$, and calculated whole-life annuity premiums, for a female age x and force of interest $\delta = 0.05$, based on each. The whole-life annuity is calculated as:

$$\overline{a}_x^g = \int_0^\infty \exp(-\delta t) \exp\left(-\int_0^t \lambda^0(x+s) RR_g ds\right) dt. \tag{6.5}$$

These premiums (relative to the baseline premium, taken to be that for $RR = 1$) were then used to construct the sampling densities of the P_g shown: (a) in Figure 6.3; and (b) singly in Figures 6.4 and 6.5, alongside the sampling distributions of the corresponding relative risks (the latter shown as the parameterised log-normal densities as opposed to the empirical densities from the simulated values).

The most dispersed sampling (premium) distribution is that for a carrier of the INS− gene ($g = 7$). This has a small sample size, 438, compared with (for ex-

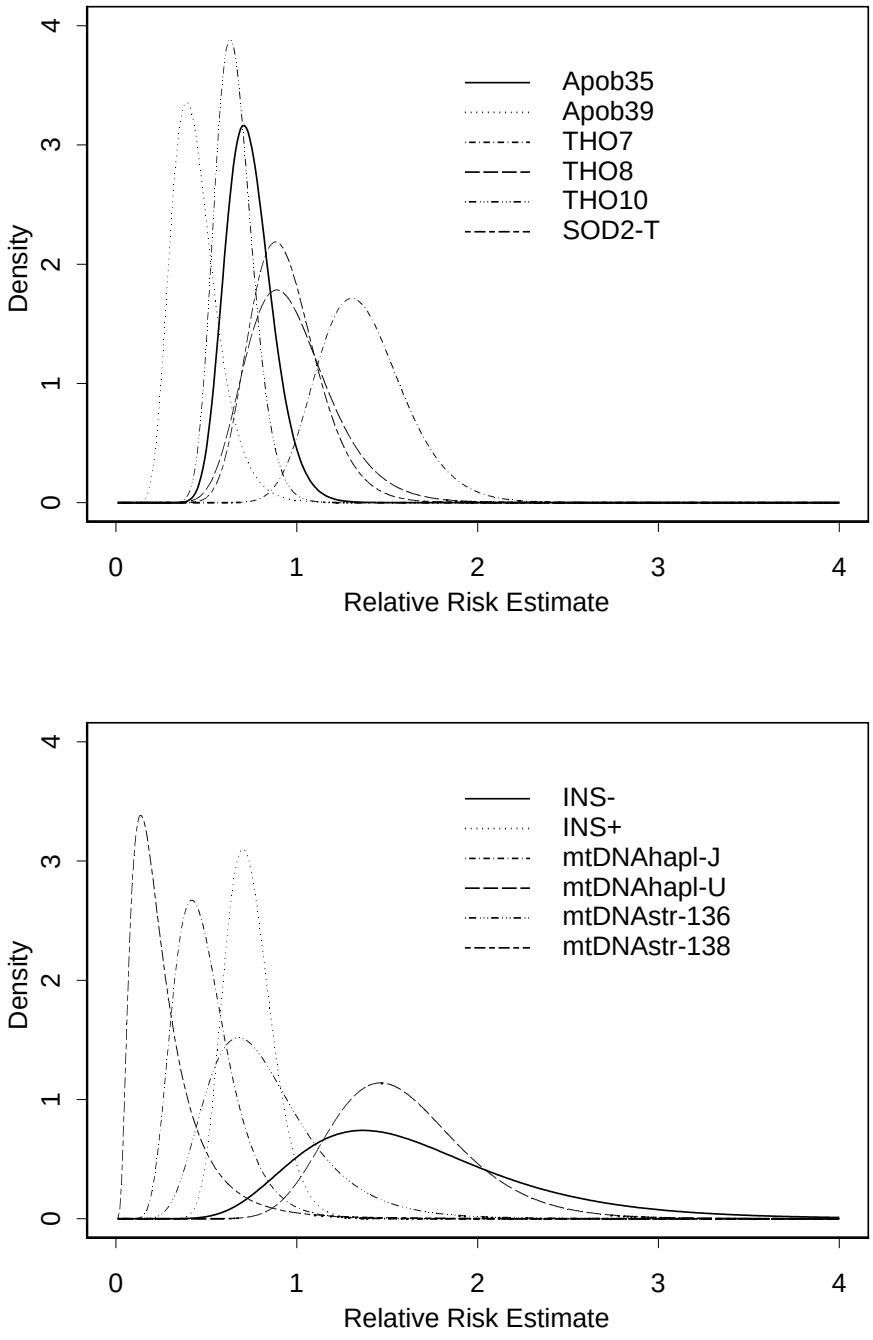


Figure 6.1: Log-normal sampling densities of the relative risk estimates for females from Tan *et al.* (2001) for genes $g = 1, \ldots 6$ (top) and $g = 7, \ldots 12$ genes (bottom).

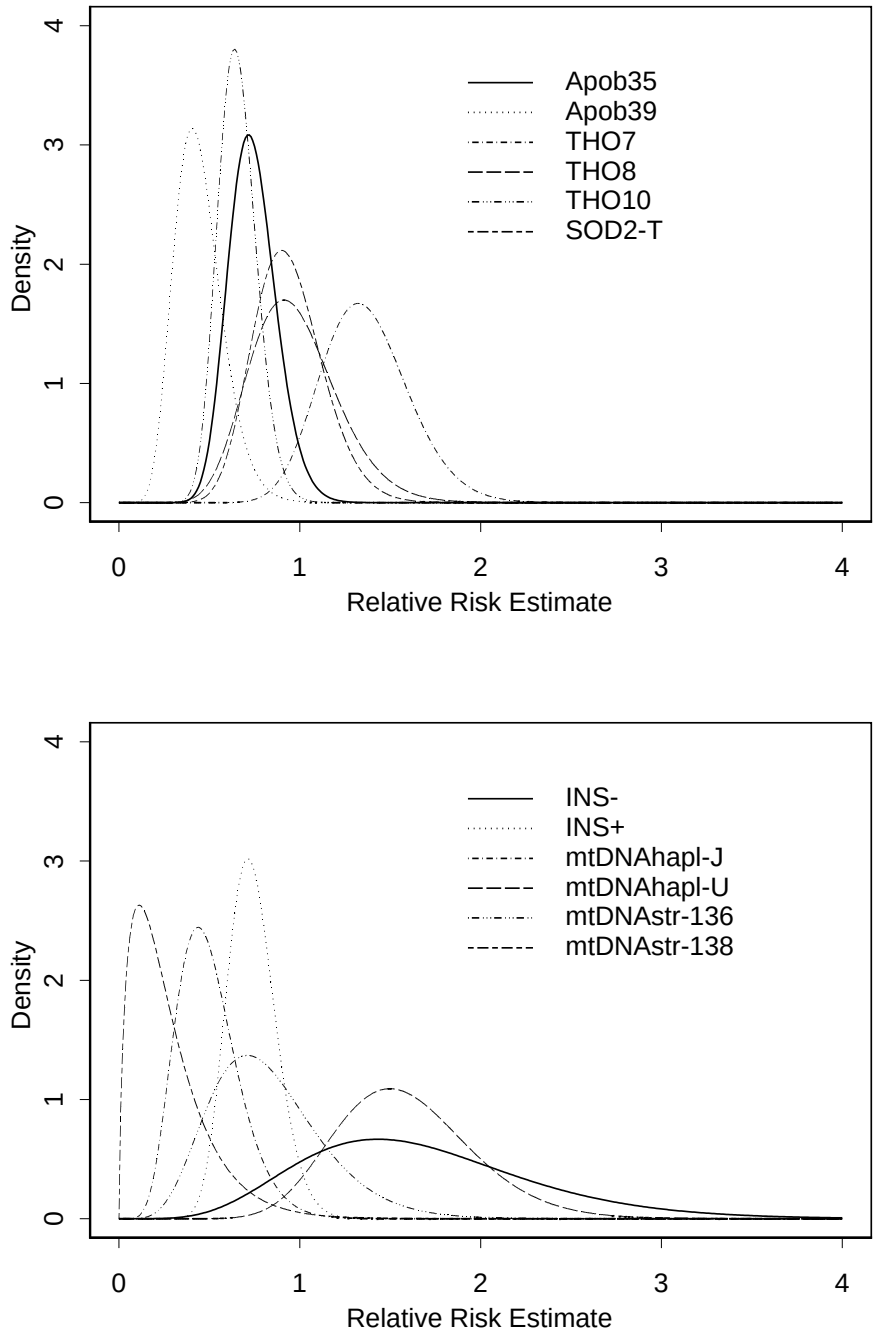


Figure 6.2: Gamma sampling densities of the relative risk estimates for females from Tan *et al.* (2001) for genes $g = 1, \ldots 6$ (top) and $g = 7, \ldots 12$ genes (bottom).

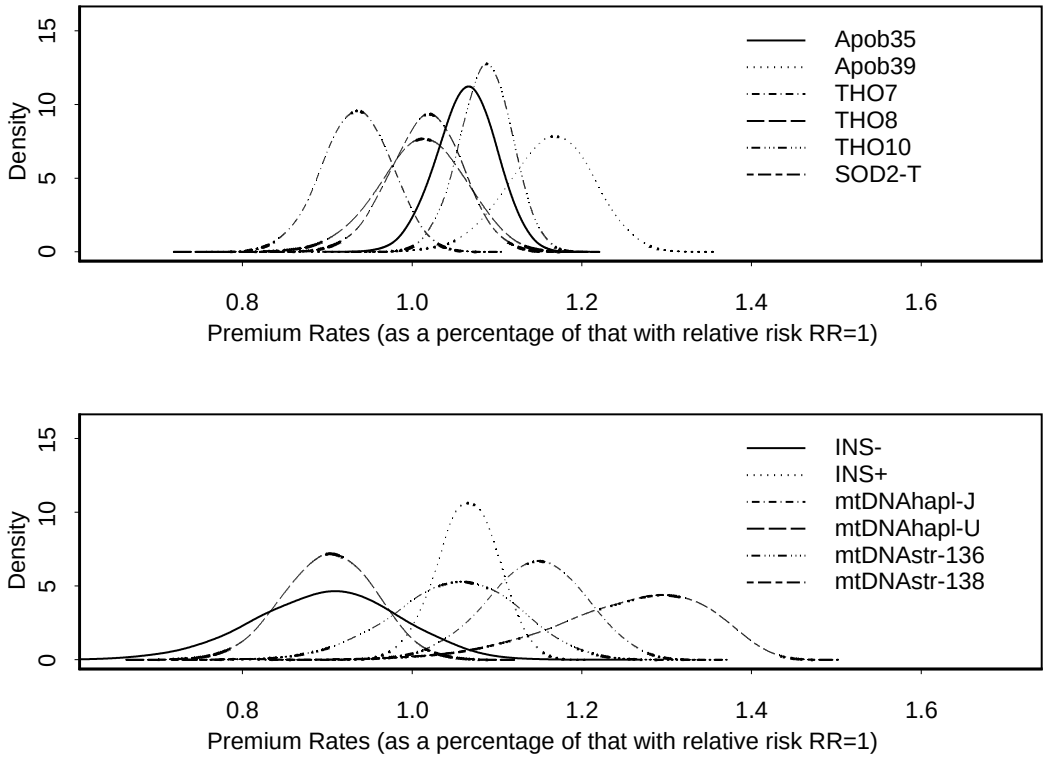


Figure 6.3: The empirical distributions of simulated single premiums for a whole-life annuity beginning at age 60 for female carriers. Genes $g = 1, \ldots 6$ are at the top and $g = 7, \ldots 12$ below.

ample) Apob35, with a sample size of 787. However the most uncertain premium estimates exist for those genes that have an extreme relative risk estimate. For example, mtDNAstr-138 has the smallest $\widehat{RR}_g$, 0.275, and its standard deviation is about average, but it produces a very dispersed premium estimate. In other words, it is not necessarily $S[\widehat{RR}_g]$ that dictates $S[\widehat{P}_g]$, but also the magnitude of $\widehat{RR}_g$.

Table 6.3 presents some key statistics of the premium sampling distributions. In particular, we show a range of sampling percentiles because they might have a rôle in deciding when a test for a particular genotype might be regarded as a relevant and reliable indicator of increased risk, given the available studies. We discuss such criteria briefly in Section 6.4.

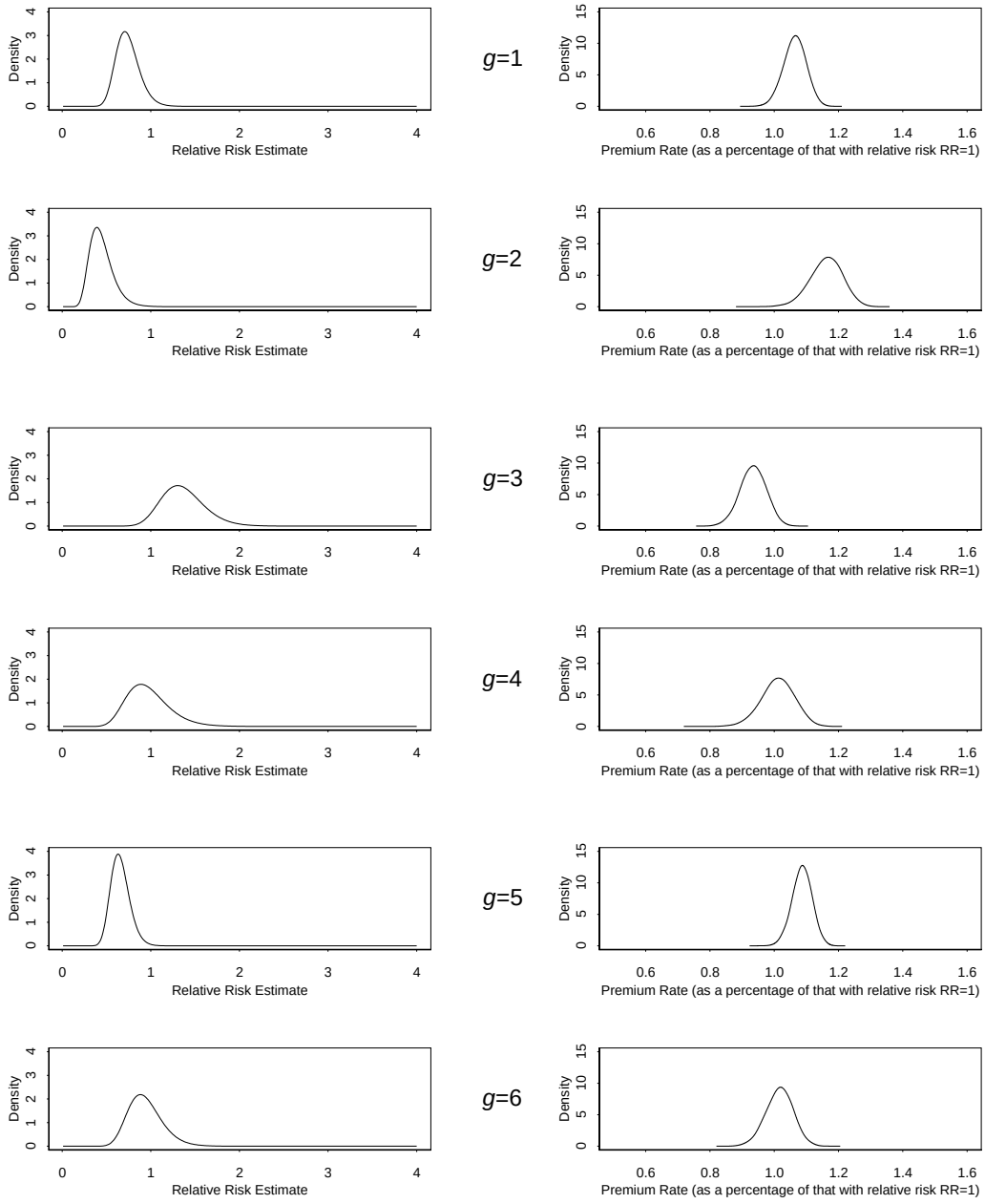


Figure 6.4: The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 1, \ldots, 6$.

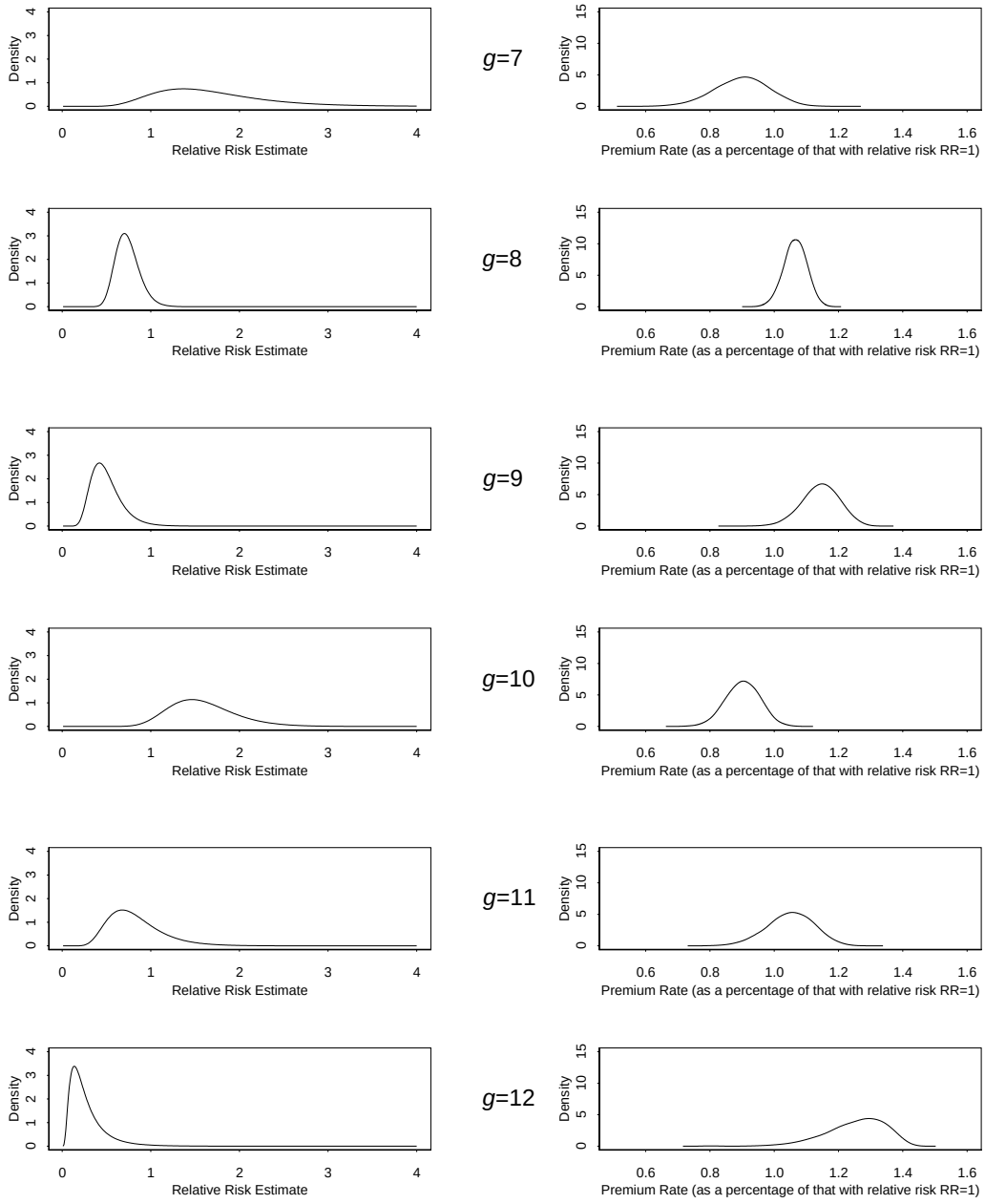


Figure 6.5: The log-normal densities of the relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for female carriers of genes $g = 7, \ldots, 12$.

Table 6.3: The mean, standard deviation and quantiles of single premiums for a whole-life annuity for a female age 60 based on a log-normal distribution of relative risk estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$.

			Quantiles of the Premium Distribution as a Percentage of the Baseline Premium								
Gene	Mean	St. Dev.	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
	%	%	%	%	%	%	%	%	%	%	%
Apob35	106.4	3.5	99.4	100.6	101.9	104.1	106.5	108.8	110.8	111.9	113.0
Apob39	116.2	4.9	106.0	107.7	109.8	113.0	116.4	119.7	122.3	124.0	125.2
THO7	93.4	4.0	85.3	86.7	88.3	90.8	93.5	96.2	98.6	99.9	101.0
THO8	101.1	5.1	90.8	92.5	94.6	97.9	101.3	104.6	107.5	109.2	110.6
THO10	108.6	3.1	102.5	103.4	104.7	106.6	108.7	110.7	112.5	113.5	114.5
SOD2-T	101.6	4.2	93.2	94.6	96.1	98.9	101.7	104.5	106.9	108.3	109.6
INS-	89.9	8.4	72.7	75.6	79.0	84.2	90.2	95.7	100.5	103.3	105.4
INS+	106.5	3.6	99.4	100.5	101.9	104.2	106.6	109.0	111.1	112.3	113.2
mtDNAhapl-J	114.4	5.8	102.6	104.5	106.9	110.7	114.6	118.5	121.7	123.6	125.1
mtDNAhapl-U	90.2	5.3	79.7	81.5	83.3	86.6	90.3	93.9	97.0	98.6	100.3
mtDNAstr-136	105.0	7.3	89.8	92.5	95.3	100.3	105.3	110.1	114.1	116.5	118.5
mtDNAstr-138	125.3	9.4	104.2	108.3	112.7	119.8	126.7	132.3	136.1	137.9	139.3

6.2.6 Relative Risks and Premiums for Males

Tan *et al.* (2001) defined relative risks for males with respect to females, requiring a product of two relative risks to be applied to the baseline hazard. This is equivalent to a gene × sex interaction term being introduced for each gene in the model. For example, suppose gene g reduces female risk and has $\widehat{RR}_g = 0.5$, but has no effect on males. Then the relative risk introduced by the gene × sex term, denoted $\widehat{RR}_g^{g\times s}$, would be 2, so that overall the relative risk for males is $\widehat{RR}_g \times \widehat{RR}_g^{g\times s} = 0.5 \times 2 = 1$.

If we assume: (a) that the sampling distributions of the $\widehat{RR}_g^{g\times s}$ are also log-normal; and (b) that the estimates $\widehat{RR}_g$ and $\widehat{RR}_g^{g\times s}$ are independent, given the data, then we can easily find the sampling distributions of $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$, since the product of log-normal(μ_1, σ_1^2) and log-normal(μ_2, σ_2^2) random variables is log-normal($\mu_1 + \mu_2$, $\sigma_1^2 + \sigma_2^2$). Assumption (a) may be reasonable, but independence is very unlikely, since given the sampling distribution of the overall relative risk for males, any shift in the marginal sampling distribution of $\widehat{RR}_g$ is likely to be compensated for by a shift in the opposite direction in the marginal sampling distribution of $\widehat{RR}_g^{g\times s}$. However, there is nothing we can do about this, lacking the sampling covariances of $\widehat{RR}_g$ and $\widehat{RR}_g^{g\times s}$. We can only comment that the results for males are more tentative than those for females.

In Figures 6.6 and 6.7 we show the sampling distributions of the relative risk estimates (parameterised densities, not simulated) alongside those of the single premium for a male age 60 (simulated) based on force of interest $\delta = 0.05$. The premium sampling distributions are generally more dispersed than those for females. This is consistent with the fact that there were fewer male centenarians in the study.

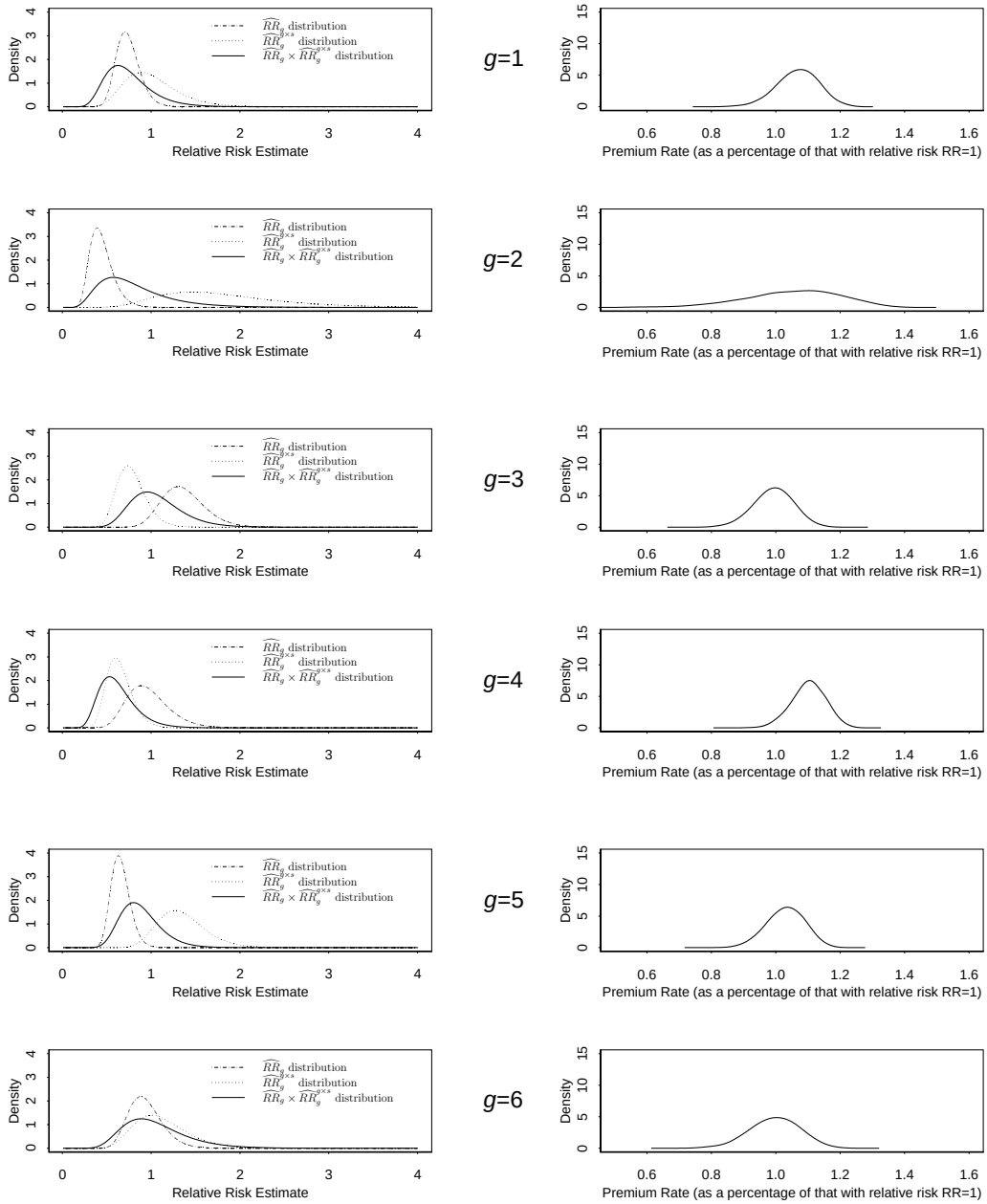


Figure 6.6: The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 1, \ldots, 6$.

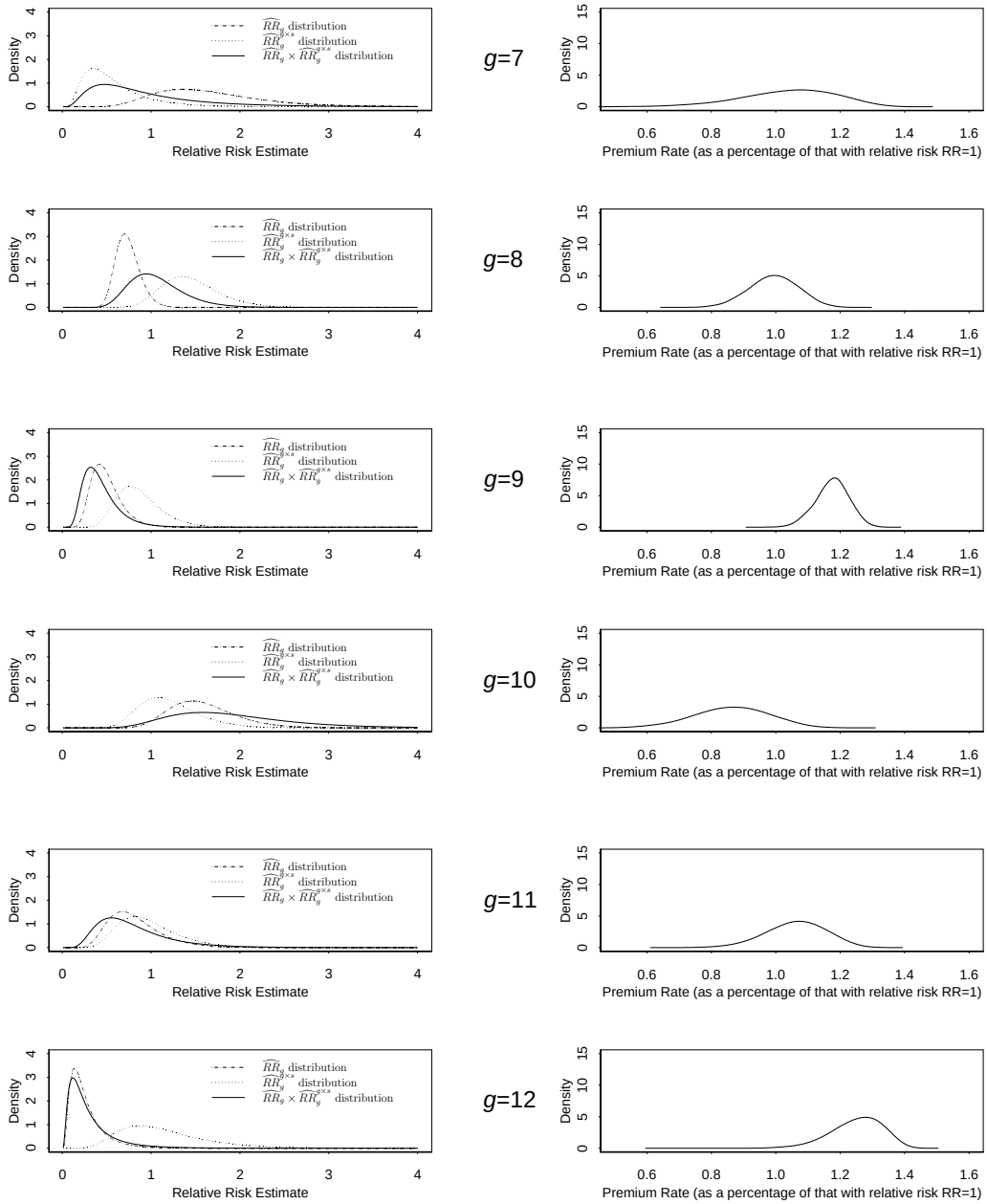


Figure 6.7: The density curves of log-normally distributed relative risk estimates $\widehat{RR}_g$, $\widehat{RR}_g^{g\times s}$ and $\widehat{RR}_g \times \widehat{RR}_g^{g\times s}$ (left) and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for male carriers of genes $g = 7, \ldots, 12$.

Table 6.4: The mean, standard deviation and quantiles of single premiums for a whole-life annuity for a male age 60 based on a log-normal distribution of relative risk estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$.

			Quantiles of the Premium Distribution as a Percentage of the Baseline Premium								
Gene	Mean	St. Dev.	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
	%	%	%	%	%	%	%	%	%	%	%
Apob35	106.8	6.5	93.4	95.5	98.4	102.5	107.1	111.4	115.0	117.0	119.0
Apob39	105.2	14.7	73.6	79.1	85.6	96.0	106.5	115.9	123.4	127.5	130.0
THO7	99.3	6.2	86.6	88.8	91.3	95.3	99.5	103.6	107.1	109.2	110.9
THO8	110.2	5.3	99.4	101.1	103.2	106.8	110.4	113.9	116.8	118.5	120.0
THO10	103.1	5.9	91.1	93.1	95.3	99.1	103.3	107.3	110.5	112.3	113.9
SOD2-T	99.6	7.8	83.4	86.6	89.5	94.5	99.8	105.1	109.5	112.1	114.1
INS-	104.0	14.7	71.7	77.6	84.3	94.6	105.3	114.7	122.1	125.8	128.7
INS+	99.4	7.6	84.3	86.7	89.4	94.3	99.5	104.6	109.0	111.4	113.6
mtDNAhapl-J	117.5	5.1	106.9	108.6	110.7	114.3	117.8	121.0	124.0	125.6	127.0
mtDNAhapl-U	86.2	11.6	62.4	66.3	71.2	78.5	86.5	94.4	101.2	104.8	107.8
mtDNAstr-136	106.3	9.3	86.9	90.3	94.2	100.2	106.6	112.8	117.9	120.7	123.1
mtDNAstr-138	125.0	8.0	107.1	110.8	114.5	120.1	125.9	130.8	134.4	136.1	137.6

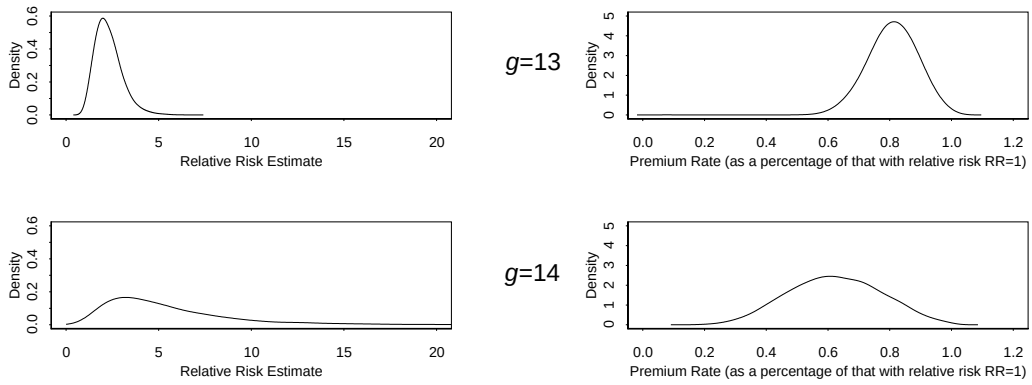


Figure 6.8: The density curves of log-normally distributed relative risk estimates (left), and the empirical densities of single premiums (right) for a whole-life annuity beginning at age 60 for carriers of genes $g = 13, 14$.

6.2.7 Relative Risks and Premiums Based on the Ashkenazi Jewish Cohort

Arking *et al.* (2005) presented parameter estimates in the form of the Cox regression coefficients rather relative risks. This is equivalent to log-normally distributed relative risks, as before.

The sampling distributions of the relative risks and (by simulation) of annuity premiums are shown in Figure 6.8 and significant quantiles are tabulated in Table 6.5. Both KLOTHO genotypes are detrimental to survival (relative to the FV genotype) and therefore significantly reduce premium rate estimates. As we would expect from such a small study (216 participants), the standard deviations of the premium rate estimates are high. The probability density of premiums remains below 100% of the baseline premium for all simulated rates (the highest for both genotypes is approximately 100.0% of the baseline, by coincidence, which is not entirely clear from Figure 6.8 because of smoothing). However, these relative risks have to be interpreted with caution. Being based on individuals over age 95, it is possible that the detrimental effects may be limited to very elderly populations.

Table 6.5: The mean, standard deviation and quantiles of single premiums for a whole-life annuity for individuals age 60 with KLOTHO genotypes FF and VV based on a Normal distribution of β estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative risk $RR = 1$.

			Quantiles of the Premium Distribution as a Percentage of the Baseline Premium								
Gene	Mean	St. Dev.	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
	%	%	%	%	%	%	%	%	%	%	%
KLOTHO FF	81.0	7.9	65.1	67.6	70.8	75.9	81.2	86.5	90.9	93.3	95.4
KLOTHO VV	62.3	14.9	33.8	38.0	42.7	51.5	62.2	73.0	82.3	87.0	91.0

Table 6.6: The APOE genotypes studied in Hayden *et al.* (2005).

Gene	Sample Size	Relative Odds Female	Male
APOE $\epsilon 2/\epsilon 2$	35	1.67	2.36
APOE $\epsilon 2/\epsilon 3$	610	1.09	0.72
APOE $\epsilon 2/\epsilon 4$	170	0.66	1.05
APOE $\epsilon 3/\epsilon 4$	1,217	1.47	1.10
APOE $\epsilon 4/\epsilon 4$	135	2.52	1.52

6.3 The APOE Genotype and Longevity

6.3.1 The APOE Genotype and Mortality

The Apolipoprotein E (APOE) gene is well-known to influence longevity. The gene can take one of three forms (alleles): $\epsilon 2$, $\epsilon 3$ or $\epsilon 4$, hence the genotype has six variants. Hayden *et al.* (2005) applied a logistic regression model to a large cohort of subjects (all over the age of 65 and followed up for seven years) and estimated relative odds of death (and their standard deviations) for APOE genotypes. Table 6.6 summarises the study, showing the relative odds statistics that we will apply in our premium calculations.

6.3.2 Logistic Regression of Survival Data

Although proportional hazards models are used widely in survival analysis, the logistic regression model often appears when dealing with interval-grouped event times (or 'tied' data) or when the proportional hazards assumption is not correct. Logistic regression coefficients do not describe the intuitive 'relative risk', but the 'relative *odds*' (also known as the 'odds ratio').

Interpretation of the relative odds is not as straightforward as that of the relative risk. The proportional hazards regression model (Cox model, see Section 6.2.1) takes the form:

$$\lambda^i(t) = \lambda^0(t) \ \exp\left(\sum_{k=1}^{p} \beta_k \ z_{ik}\right) \tag{6.6}$$

while the logistic regression model is of the form:

$$\frac{\lambda^i(t)}{1-\lambda^i(t)} = \frac{\lambda^0(t)}{1-\lambda^0(t)} \ \exp\left(\sum_{k=1}^{p} \beta_k \ z_{ik}\right). \tag{6.7}$$

In both models the relation between the hazards is given by a regression component of the form $\exp(\sum_{k=1}^{p} \beta_k \ z_{ik})$, but the estimated beta coefficients in each case are different.

We can express relative risks in terms of constant relative odds. For brevity, let $\exp(\sum_{k=1}^{p} \beta_k \ z_{ik})$ in Equation (6.7) be denoted RO_i. From Equation (6.7):

$$\begin{aligned} \frac{\lambda^i(t)}{\lambda^0(t)} &= \frac{1}{\lambda^0(t)} \times \frac{\frac{\lambda^0(t)}{1-\lambda^0(t)} \ RO_i}{1+\frac{\lambda^0(t)}{1-\lambda^0(t)} \ RO_i} && (6.8) \\ &= \frac{RO_i}{1-\lambda^0(t)+\lambda^0(t) \ RO_i}. && (6.9) \end{aligned}$$

However, the relative risk is no longer a constant, but a function of t, that we may denote $RR_i(t)$. We can see how the relative risk changes for different baseline hazard rates, with respect to different relative odds, in Figure 6.9. This suggests that approximating relative risks by relative odds in studies of longevity, where baseline mortality at older ages can be high, could yield misleading conclusions. Of course, if we have a good estimate of $\lambda^0(t)$, we can simply apply Equation (6.9) directly. We can use this to compute any relevant actuarial quantities. It is now the relative odds RO_i that is assumed to have an approximately log-normal sampling distribution, from which we can simulate values and, via Equation (6.9), estimate the sampling distributions of any derived actuarial quantities.

Now that we have shown that the relative risk is related to the relative odds by the function given in Equation (6.9), it is interesting to observe how the distribution of relative risk is affected by different values of $\lambda^0(t)$. The distribution of $RR_i(t)$ when $RO_i \sim$ log-normal(0,0.25) is shown in Figure 6.10. Note that for low values of $\lambda^0(t)$ the distribution of relative risk is approximately log-normal, and approaches a constant as $\lambda^0(t)$ increases.

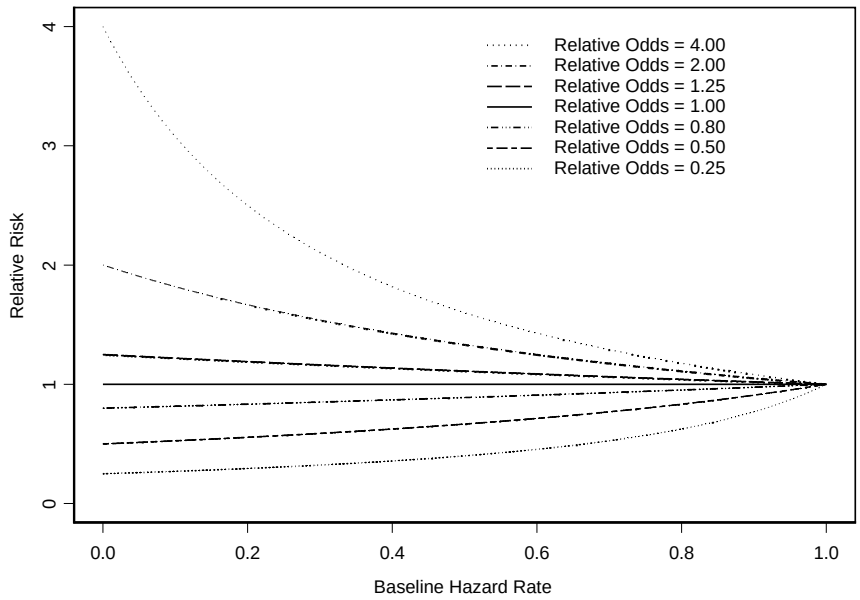


Figure 6.9: The relative risk through different values of the hazard rate $\lambda_0(t)$ calculated for several relative odds values.

6.3.3 Premium Rate Sampling Distributions Given APOE Genotype

Figure 6.11 shows the sampling distributions of premium rates (for a whole-life annuity issued to a life aged 65, with force of interest $\delta = 0.05$) for APOE genotypes $\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$, relative to the premium rates in respect of the $\epsilon 3/\epsilon 3$ genotype. These are based on 10,000 simulations from the approximately log-normal sampling distribution of the relative odds. The baseline hazard rate we used was USA mortality from calendar year 2000, as the observations were made at approximately that time. That is, we simply attribute USA population mortality to carriers of the most common genotype, $\epsilon 3/\epsilon 3$.

Table 6.7 shows the mean relative premium rates and 95% confidence intervals. The notable frailty genotypes seem to be $\epsilon 2/\epsilon 2$ and $\epsilon 2/\epsilon 3$ for females, and $\epsilon 4/\epsilon 4$ for males. The genotype $\epsilon 3/\epsilon 4$ appears to be a frailty genotype in females but a longevity genotype in males. By contrast, $\epsilon 2/\epsilon 4$ is a longevity genotype in females but a frailty

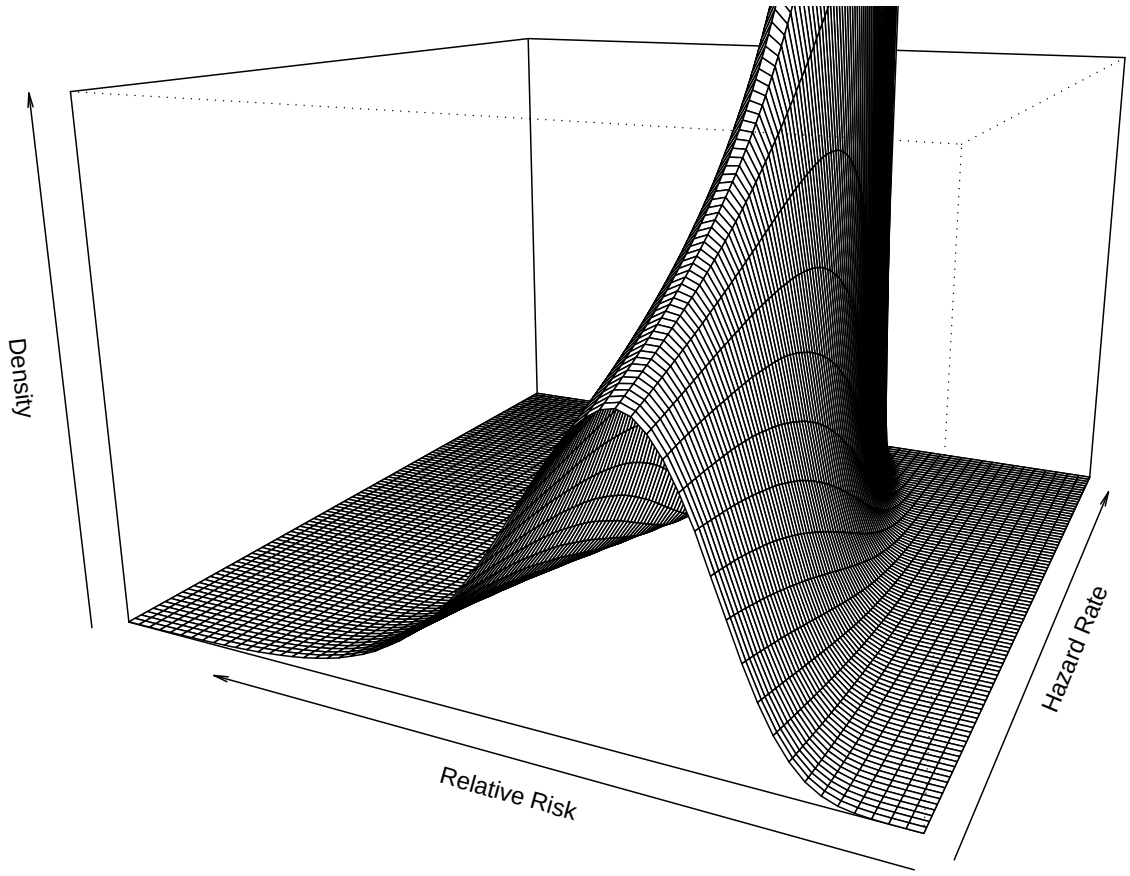
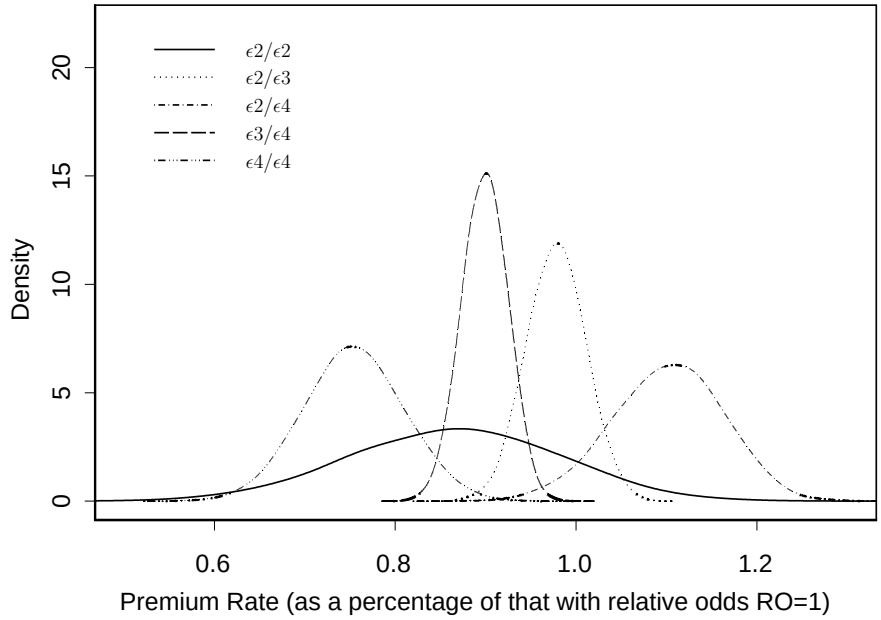


Figure 6.10: The distribution of relative risk throughout different values of the hazard rate $\lambda_0(t)$ assuming the relative odds are distributed log-normally. Graph is based on $RO_i \sim$ log-normal(0,0.25).



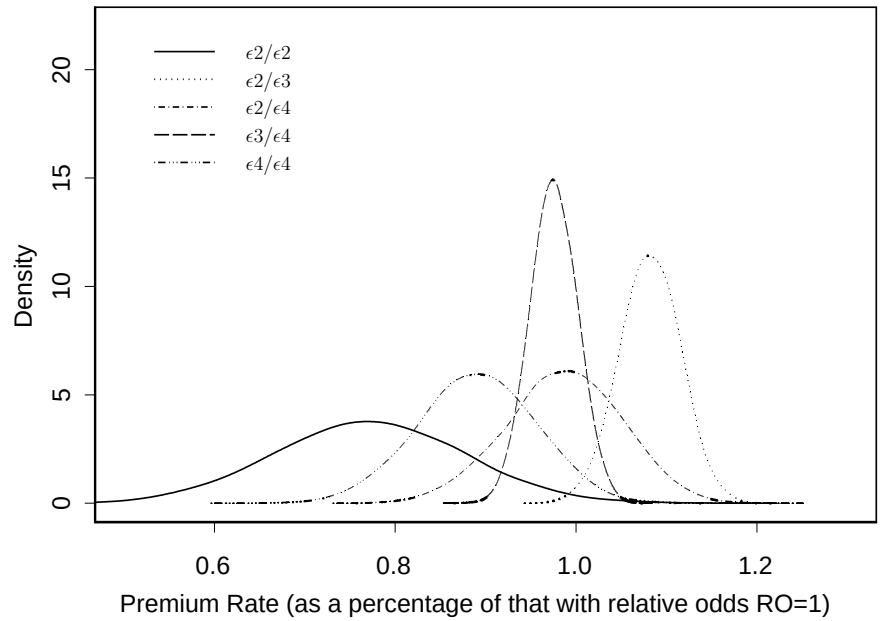


Figure 6.11: The empirical densities of whole-life annuities for a female (top) and a male (bottom) beginning at age 65, for APOE genotypes $\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 4$, and $\epsilon 4/\epsilon 4$ relative to the annuity cost of a $\epsilon 3/\epsilon 3$ genotype carrier.

genotype in males.

Table 6.7: Single premiums for level whole-life pension annuities of 1 per year payable continuously, depending on APOE genotype. The premiums are expressed as a percentage of those for the most common genotype, $\epsilon 3/\epsilon 3$. Premiums are shown for healthy male and female purchasers aged 65, 70 and 75.

		Premium as % of that for $\epsilon 3/\epsilon 3$ Genotype (95% Confidence Intervals)				
Gender	Entry Age	$\epsilon 4/\epsilon 4$	$\epsilon 3/\epsilon 4$	$\epsilon 2/\epsilon 4$	$\epsilon 2/\epsilon 3$	$\epsilon 2/\epsilon 2$
Female	65	75.37	89.90	110.28	97.78	86.48
		(64.65, 86.29)	(84.97, 94.79)	(97.80, 121.49)	(91.27, 104.03)	(63.72, 108.58)
	70	72.36	88.36	112.27	97.41	84.52
		(61.25, 84.00)	(82.82, 93.92)	(97.29, 126.30)	(90.00, 104.77)	(60.04, 110.14)
	75	69.57	86.81	114.49	97.01	82.57
		(58.13, 82.19)	(80.78, 93.05)	(96.84, 131.71)	(88.58, 105.47)	(57.17, 112.08)
Male	65	89.01	97.54	98.75	108.19	77.13
		(76.26, 101.34)	(92.47, 102.54)	(85.78, 110.81)	(101.56, 114.46)	(57.34, 97.48)
	70	87.35	97.13	98.53	109.74	74.26
		(73.07, 101.67)	(91.23, 103.04)	(83.84, 112.90)	(101.86, 117.49)	(53.71, 96.83)
	75	85.69	96.70	98.30	111.46	71.56
		(70.76, 102.00)	(90.01, 103.44)	(81.91, 115.45)	(102.01, 120.71)	(50.84, 96.69)

Table 6.8: The mean, standard deviation and quantiles of single premiums for a whole-life annuity for males and females age 65 based on a log-normal distribution of relative odd estimates. They are expressed as percentages of a baseline premium rate, taken to be that for relative odds $RO = 1$.

				Quantiles of the Premium Distribution as a Percentage of the Baseline Premium								
Gender	Genotype	Mean %	St. Dev. %	2.5th %	5th %	10th %	25th %	50th %	75th %	90th %	95th %	97.5th %
Female	$\epsilon4/\epsilon4$	75.4	5.5	64.6	66.4	68.4	71.7	75.4	79.2	82.5	84.6	86.3
	$\epsilon3/\epsilon4$	89.9	2.5	85.0	85.8	86.7	88.2	89.9	91.6	93.1	94.0	94.8
	$\epsilon2/\epsilon4$	110.1	6.1	97.8	99.9	102.3	106.2	110.3	114.3	117.7	119.7	121.5
	$\epsilon2/\epsilon3$	97.8	3.2	91.3	92.4	93.6	95.6	97.8	100.0	102.0	103.1	104.0
	$\epsilon2/\epsilon2$	86.5	11.5	63.7	67.2	71.5	78.5	86.5	94.4	101.4	105.1	108.6
Male	$\epsilon4/\epsilon4$	89.0	6.4	76.3	78.4	80.8	84.7	89.1	93.3	97.2	99.4	101.3
	$\epsilon3/\epsilon4$	97.5	2.6	92.5	93.2	94.1	95.8	97.5	99.3	100.8	101.8	102.5
	$\epsilon2/\epsilon4$	98.7	6.3	85.8	87.9	90.2	94.3	98.8	103.0	106.7	108.8	110.8
	$\epsilon2/\epsilon3$	108.2	3.3	101.6	102.6	103.9	105.9	108.2	110.4	112.3	113.5	114.5
	$\epsilon2/\epsilon2$	77.1	10.3	57.3	60.6	64.0	70.2	77.1	84.3	90.7	94.3	97.5

Table 6.9: Single premiums for level whole-life pension annuities of 1 per year payable continuously based on the Alzheimer's disease model of Macdonald & Pritchard (2001), treating APOE genotypes as underwriting classes. The premiums are expressed as a percentage of those for the most common genotype, $\epsilon 3/\epsilon 3$. Premiums are shown for healthy male and female purchasers aged 60, 65, 70 and 75.

		Premium as % of that for $\epsilon 3/\epsilon 3$ Genotype			
Gender	Entry Age	$\epsilon 4/\epsilon 4$	$\epsilon 3/\epsilon 4$	$\epsilon 2/\epsilon 4$	$\epsilon 2/\epsilon 2$ & $\epsilon 2/\epsilon 3$
Female	60	95.47	97.79	97.61	100.33
	65	95.45	97.49	97.07	100.43
	70	96.33	97.59	97.31	100.51
	75	97.77	98.13	98.35	100.50
Male	60	96.36	99.70	99.78	100.43
	65	96.78	99.72	99.70	100.47
	70	97.93	99.78	99.70	100.41
	75	99.16	99.77	99.63	100.12

The rate of AD onset is well-known to be raised by the $\epsilon 4$ APOE allele, and is thought to be lowered by the $\epsilon 2$ allele. Macdonald & Pritchard (2000, 2001) used the meta-analysis by Farrer *et al.* (1997) to parameterise a model for long-term care (LTC) allowing for AD onset, and subsequent institutionalisation and/or mortality, and used it to price a combined LTC insurance and pension package. To compare this model with ours, we have used it to price a level, continuous annuity of 1 per annum payable while alive, with no LTC benefit (Table 6.9). Their model was restricted in the sense that the APOE gene influenced death through AD alone. Our work takes into account the influence of APOE genotype on other causes of death (cardiovascular disease for example) so our results differ considerably from theirs. Bearing these disparities in mind, we can compare the first three columns of Tables 6.7 and 6.9. With the exception of the $\epsilon 4/\epsilon 4$ genotype, the figures from Table 6.9 are sometimes close to the 95% confidence limits in Table 6.7. The similarities suggest that death from APOE-related AD is an important cause of mortality at late ages, but other significant APOE-related causes of mortality are apparent, and strongly linked with the $\epsilon 4/\epsilon 4$ genotype.

6.4 Discussion and Conclusions

6.4.1 Acceptable Uncertainty

Relative risk estimates are often published in epidemiological studies. We have highlighted their sampling properties and the sampling distributions that are inherited by premium rates based on them. Many of the genes in this study might at first sight appear to be financially important; however the sampling distributions of the corresponding premium rates introduce much more uncertainty. This kind of statistical information is relevant to any consideration of using genotype information in insurance practice, for example in the deliberations of the Genetics and Insurance Committee (GAIC) in the UK.

GAIC was charged, by the UK Government, with ensuring that any use of genetic test results by insurers would have a sound actuarial and scientific basis. To date, GAIC has approved only one genetic test (for Huntington's disease, in the case of life insurance) as a reliable predictor of significantly increased risk. GAIC will face difficult questions if it is required to review tests for more complex disorders, whose results do not indicate an overwhelming increase or decrease in population mortality, and as such sampling error should be taken into account. This is the case for longevity genes. When presented with sampling distributions of relative premium rates, GAIC will have to answer questions such as: what percentile of the premium rate sampling distribution might justify the use of a test by insurers? (we call this an 'acceptance percentile', but it is only one of many criteria that could be adopted). Such questions may well arise as research into genes with modest effects on mortality and morbidity (by current standards) enters the medical mainstream.

If an acceptance percentile were adopted as a criterion for the use of a genetic test, the next question is: what premium loading should be applied? This question is probably beyond GAIC's remit, and would be left to individual insurers, who would be allowed to take into consideration their different risk tolerances and underwriting practices.

6.4.2 Acceptance Percentiles

As an example of how 'acceptance percentile' might contribute to such decisions as described above, Table 6.10 shows which genotypes might be regarded as having a significant impact at a 75%, 90%, 95% and 97.5% level, based on percentiles from Tables 6.3, 6.4, 6.5 and 6.8. Note that upper or lower percentiles are used, depending on whether a gene is a candidate longevity gene or a candidate frailty gene.

If the criterion of a one-tailed 97.5% confidence interval (of annuity prices) were adopted (implying very low uncertainty) then, among the genes considered by Tan *et al.* (2001), for females, four 'longevity' gene variants would appear to be important: Apob39, THO10, mtDNAhapl-J and mtDNAstr-138. For males, there would be only two: mtDNAhapl-J and mtDNAstr-138. Tan *et al.* (2001) estimated the frequency of these genes in the Italian population — Apob39 and THO10 are both common (frequency $\approx$ 30–40%) whereas mtDNAhapl-J and mtDNAstr-138 are relatively rare (frequency $\approx$ 1–5%). Therefore, in the scenario of widespread genetic testing the genes Apob39 and THO10 could lead to large-scale segmentation of the annuity market (always supposing that results based on an Italian population generalise to other populations).

The APOE genotype is arguably more important. Commercial testing for APOE genotype is readily available and Hayden *et al.* (2005) is only one of many research teams that have confirmed its rôle in longevity. Most APOE genotypes are frailty genotypes that act to reduce annuity premiums (relative to the $\epsilon 3/\epsilon 3$ norm).

Our methodology is not confined to genetic risk. Indeed we are surprised that it has taken genetic risk to draw attention to sampling issues in actuarial estimates based on epidemiological and medical studies. Consideration of premium rate sampling error would seem to be an elementary extension from professional statistical practice to professional actuarial practice.

Table 6.10: A list of all genes/genotypes studied, and whether they are significant at a 75%, 90%, 95% or 97.5% level. A ✓ represents a significant gene/genotype and a ✗ represents a non-significant gene/genotype. The phenotype is the observable manifestation of the gene/genotype, this is either frailty or longevity.

Gender	Gene/Genotype	75%	90%	95%	97.5%	Phenotype
Female	Apob35	✓	✓	✓	✗	Longevity
	Apob39	✓	✓	✓	✓	Longevity
	THO7	✓	✓	✓	✗	Frailty
	THO8	✗	✗	✗	✗	Longevity
	THO10	✓	✓	✓	✓	Longevity
	SOD2-T	✗	✗	✗	✗	Longevity
	INS-	✓	✗	✗	✗	Frailty
	INS+	✓	✓	✓	✗	Longevity
	mtDNAhapl-J	✓	✓	✓	✓	Longevity
	mtDNAhapl-U	✓	✓	✓	✗	Frailty
	mtDNAstr-136	✓	✗	✗	✗	Longevity
	mtDNAstr-138	✓	✓	✓	✓	Longevity
	APOE $\epsilon 4/\epsilon 4$	✓	✓	✓	✓	Frailty
	APOE $\epsilon 3/\epsilon 4$	✓	✓	✓	✓	Frailty
	APOE $\epsilon 2/\epsilon 4$	✓	✓	✗	✗	Longevity
	APOE $\epsilon 2/\epsilon 3$	✓	✗	✗	✗	Frailty
	APOE $\epsilon 2/\epsilon 2$	✓	✗	✗	✗	Frailty
Male	Apob35	✓	✗	✗	✗	Longevity
	Apob39	✗	✗	✗	✗	Longevity
	THO7	✗	✗	✗	✗	Frailty
	THO8	✓	✓	✓	✗	Longevity
	THO10	✗	✗	✗	✗	Longevity
	SOD2-T	✗	✗	✗	✗	Frailty
	INS-	✗	✗	✗	✗	Longevity
	INS+	✗	✗	✗	✗	Frailty
	mtDNAhapl-J	✓	✓	✓	✓	Longevity
	mtDNAhapl-U	✓	✗	✗	✗	Frailty
	mtDNAstr-136	✓	✗	✗	✗	Longevity
	mtDNAstr-138	✓	✓	✓	✓	Longevity
	APOE $\epsilon 4/\epsilon 4$	✓	✓	✓	✗	Frailty
	APOE $\epsilon 3/\epsilon 4$	✓	✗	✗	✗	Frailty
	APOE $\epsilon 2/\epsilon 4$	✗	✗	✗	✗	Frailty
	APOE $\epsilon 2/\epsilon 3$	✓	✓	✓	✓	Longevity
	APOE $\epsilon 2/\epsilon 2$	✓	✓	✓	✓	Frailty
Both	KLOTHO FF	✓	✓	✓	✓	Frailty
	KLOTHO VV	✓	✓	✓	✓	Frailty

Chapter 7

Conclusions and Further Work

7.1 Conclusions

7.1.1 The Polygenic Model

The polygenic approach to studying disease risk is enabled by the vast information on genetic variation that is being harvested from the human genome. It is generally accepted that most common disorders, such as cancers and heart disorders, have a genetic component which is related to the status of several, perhaps common, gene variants. As a means of studying disease pathology, the polygenic model, in some form, is likely to appear widely in the future of genetic studies.

We have used a polygenic model fitted to a UK population of families with breast and ovarian cancer to price CI insurance policies with varying terms and entry ages. Then, by estimating the genotype frequencies in individuals who have a family history of BC or OC, we price a CI policy that would charge the 'actuarially fair' price to those with such a history.

We considered the problems posed by adverse selection using both multi-state Markov models and utility theory models. This allowed us to take two complimentary perspectives on the problem. Firstly, multi-state modelling gave us the opportunity to draw results on several scenarios of testing and insurance buying behaviour. And secondly, the use of utility models gave us insight into the proportion of a risk-

averse population that would forego insurance provision on the basis of their utility expectations.

We reiterate that our analysis of the impact of testing for the BC polygene should be considered in partnership with the fact that polygene testing is not presently available. We lack the evidence to affirm whether the number of loci in the polygene is closer to three than it is to thirty. Indeed, the polygenic model adopted by Antoniou *et al.* (2002) may have misjudged the true synergies between the genes that comprise the polygene.

What we can say with certainty is that a polygene for BC does exist and progress towards finding the genes that compose this is being made. New technological advances have presented the ability to analyse hundreds of thousands of single nucleotide polymorphisms (SNPs) in association studies. This enables the discovery of genes which confer moderate risk of a disease without first estimating the approximate location of the gene (usually achieved by conducting a linkage analysis).

A large, world-wide association study by Easton *et al.* (2007) identified four genes with common alleles, and an unidentified gene in a known region, that contribute to the risk of developing BC. All four genes had never before been studied for association with BC mainly because they are not involved in DNA repair or related to sex hormones. The most significant SNP was found in the gene FGFR2. This SNP, in its homozygous, risk-conferring form, is present within approximately 14% of the UK population and 19% of UK BC cases. Easton *et al.* (2007) estimate a 10.5% risk of BC by age 70 for the risk-conferring, homozygous genotype, compared to 5.5% for the more common homozygous genotype. The susceptibility loci that the authors found explain a considerable portion of the genetic influence on BC, yet they believe that further such studies, with larger study sizes, should take place before testing for combinations of common, low-risk genes becomes clinically important.

Future developments in the search for the polygene could lead to improvements in population-based screening programmes and greater participation in such programmes. Some of the leading researchers in the search for the BC polygene believe that the ability to identify those at highest risk and offer early clinical intervention

will be a large incentive for those individuals to become tested (Bradbury, 2002). If all BC susceptibility genes are found, the implications for insurance will largely depend on the numbers who submit to testing and the effectiveness of available treatment.

7.1.2 Longevity

In the past few years much attention has been drawn to the issue of "the ageing population". In 2001 the number of individuals in the UK over the age of 65 became, for the first time, greater than the number under 16. This is a result of both improvements in mortality and reduced family sizes. The progressive ageing of the population in this manner places strain on health care and retirement income provision. A UK government report (Select Committee on Economic Affairs, 2004), which recognised the media's concern over the possibility of a "pensions crisis", recommended further study and oversight of the situation.

We looked at the variation in human longevity that is attributed to genetics. We searched out three genetic studies which estimated the relative risk or odds ratio, along with their standard deviations, of mortality based on an individual's carrier status for an assortment of genes. By making some assumptions about the sampling distribution of the estimates, we ascertained the distribution of annuity costs implied by the uncertainty of the estimates. This provided a method by which to judge the reliability of risk factors with respect to insurance and highlighted some genes that have considerable significance for annuity pricing.

7.2 Further Work

7.2.1 Realisation of the Polygenic Model

The gene discoveries made by Easton *et al.* (2007) represent a large step on the path of making polygene testing a reality. However, obtaining the full picture of how the polygene confers BC risk will be hindered by many difficulties that include gene-gene interaction (epistasis) and gene-environment interaction. Overcoming such

complications is liable to require much innovation on the part of geneticists and epidemiologists. One author, Tamimi (2006), believes that future studies of BC should adopt surrogate markers for the disease which are continuous (such as mammographic density) and treat each of the variety of BC diagnoses as distinct conditions.

In regards to the BRCA1/2 major genes, further understanding of their involvement with BC and OC may warrant further study of their implications for life and CI insurance. For example, no actuarial study to-date has accounted for mutation position (the BRCA2 gene alone has 500 known distinct mutations (see Table A.1)). The ovarian cancer cluster region (OCCR) is an area of the BRCA2 gene on which mutations are associated with an almost doubled lifetime risk of OC (Thompson & Easton, 2001). Mutations inside the OCCR on the BRCA2 gene may also be associated with a reduced risk of BC (Al-Saffar & Foulkes, 2002). Thus further epidemiological study may reveal risks specific to BRCA1/2 mutation position.

7.2.2 Polygenic Models in Other Diseases

The genetics and insurance debate has mainly focused, for good reasons, on monogenic disorders, a prime example being Huntington's disease. Now, genetic technologies are advancing rapidly, and we must broaden the focus to include polygenic disorders. This is a more significant undertaking than might first be thought, since every conceivable disorder can be considered to be, to some degree, polygenic. This includes the common disorders like heart disease, cancers, and autoimmune diseases. Many common diseases show familial inheritance but no single genes have been found to account for this.

Interactions between genes and environmental factors make it difficult to identify polymorphisms that influence common diseases. However, large-scale studies such as UK Biobank are now setting out to map the links between genes and environment. Medical benefits are not expected to appear for at least ten years. When results do begin to come through, however, it is likely that we will find common low-risk genes (polygenes) that are risk factors for a variety of common disorders. It is a prudent

pre-emptive step to try to understand the effect that identified polygenes may have on insurance markets.

7.2.3 Further Insurance Models

Our work regarding the polygene has considered only the implications to CI insurance. Extending this to a model of life insurance would require the transition intensities from any of the CI states into the Dead state (upon which benefits are claimed). Given that these intensities are likely to be dependent upon the time spent suffering from a CI, the multi-state model that would be needed to model life insurance would probably be semi-Markov. Also, the post-onset mortality of BRCA1/2 mutation carriers is similar to those of non-mutation carriers (Rennert *et al.*, 2007), so we could perhaps assume a common intensity as in Gui *et al.* (2006), however making the same assumption about post-onset mortality based on polygenotype status may not be viable.

The models for estimating adverse selection (namely, those in Chapters 4 and 5) address only part of the problem; in these models we assume that genetic testing is only available for genes which affect BC and OC. In the introduction we discussed several other genetic disorders for which adverse selection has been modelled. Considering each disorder individually in this way may be referred to as a 'bottom-up' approach to finding the full cost of adverse selection. This is in contrast to the more general 'top-down' approach of earlier research (Macdonald, 2000). The construction of a full model that incorporates all genetic disorders is currently underway.

We have modelled adverse selection as a greater tendency for individuals to purchase insurance on receipt of genetic test that identifies a major gene mutation or a dangerous polygenotype. However, adverse selection can also occur when tested individuals elect to purchase greater amounts of insurance than usual. It would be interesting to extend our model to consider this possibility.

Appendix A

Genes Conferring BC Risk

The genes listed alongside BRCA1 and BRCA2 in Table A.1 are candidate polygenes for BC susceptibility. A polymorphism is defined as an allele with a population frequency of at least 1% (less common alleles are more commonly referred to as 'mutations'). Polymorphisms are extremely common in the human genome (200,000–400,000; Easton, 1999) and therefore offer a vast search region for cancer susceptibility polygenes. In 2005–2006 there has been an explosion in published research related to polymorphisms associated with BC (and OC). A quick search of a medical research database (Entrez PubMed) reveals 58 papers published between 1st January 2006 and 11th May, 2006.

A number of studies have made attempts to identify existing BC susceptibility polymorphisms. Dunning *et al.* (1999) provide a summary of these studies and, from multiple independent investigations of some gene variants, conduct meta-analyses to detect significant polymorphisms that influence the risk of BC. We present the results of some of the most widely studied polymorphisms in the form of forest plots (or meta-analysis plots) in Figures A.1, A.2 and A.3. On these plots, shaded squares represent the point estimates of odds ratios (with the size of the square proportional to the significance of the estimate), shaded diamonds represent the odds ratio statistics (including 95% confidence levels) found from joint analysis by Dunning *et al.* (1999), and horizontal bars represent 95% confidence levels (clipped at an odds ratio of 5) for the odds ratios.

Table A.1: List of genes which may confer additional BC risk, Rebbeck *et al.* (1999), Easton *et al.* (1999). The allele frequencies are for possible risk-conferring polymorphisms estimated from healthy Caucasian control populations and the numbers of distinct mutations are taken from the Human Gene Mutation Database.

Gene	Allele Frequency	No. of Mutations	BC Risk
BRCA1	0.051%	741	High
BRCA2	0.068%	500	High
TP53	39%	139	High
PTEN	<0.01%	170	High
MSH2		337	High
ATM	1%	421	Moderate
CYP1A1	3-11%	2	Moderate
CYP2D6	9%	30	Low
CYP2E1	7-9%	2	Low
CHEK2	1.1%	23	Low
GSTM1	38-62%	3	Low
HRAS1	6%	1	Low
NAT2	56-62%	9	Low

Upon analysis, Dunning *et al.* (1999) identified the genes CYP19, GSTM1, GSTP1 and TP53 as candidates for low-penetrance BC susceptibility genes. We can see from the joint analyses in Figures A.1, A.2 and A.3 that polymorphisms within these genes are significant (or almost significant) at the 95% confidence level.

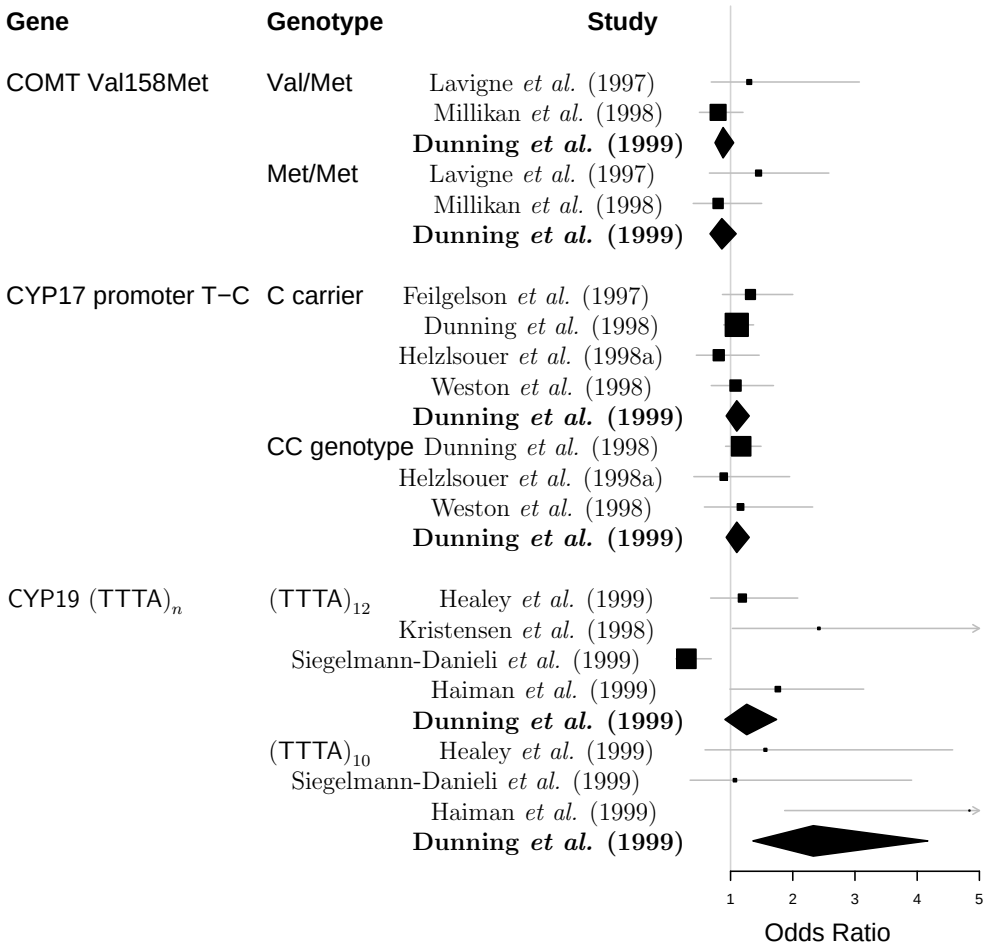


Figure A.1: Forest plot of odds ratio estimates for the genes COMT, CYP17 and CYP19, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals.

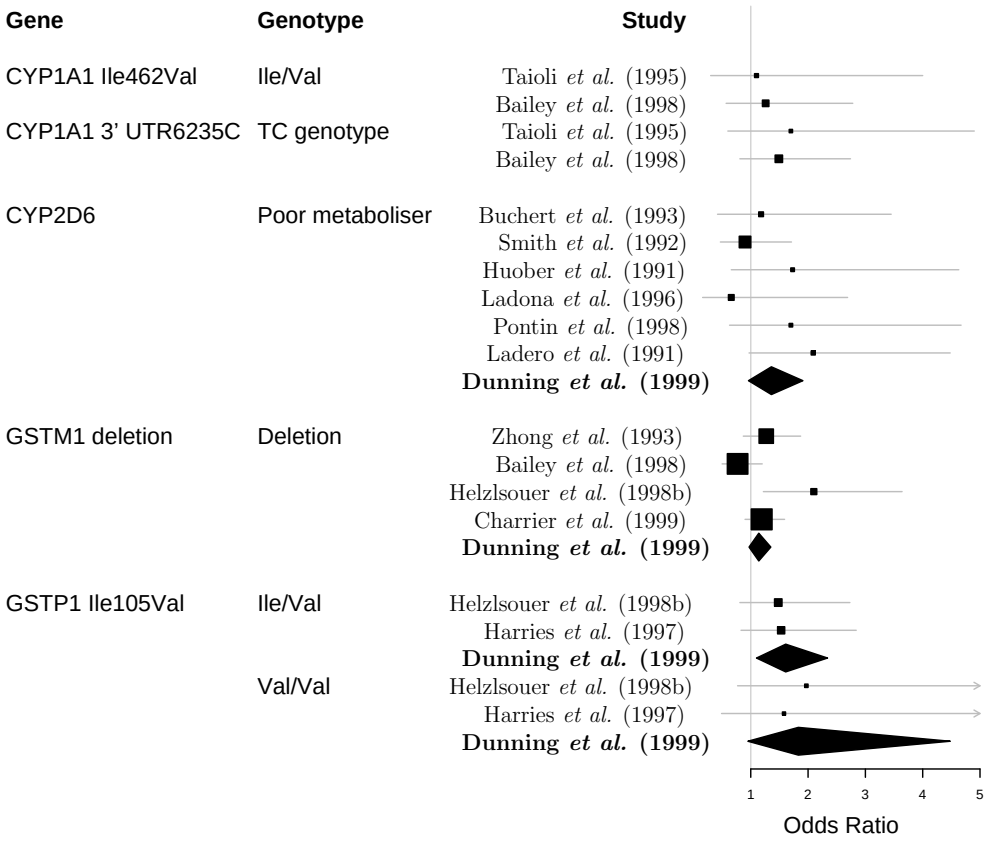


Figure A.2: Forest plot of odds ratio estimates for the genes CYP1A1, CYP2D6, GSTM1 and GSTP1, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals.

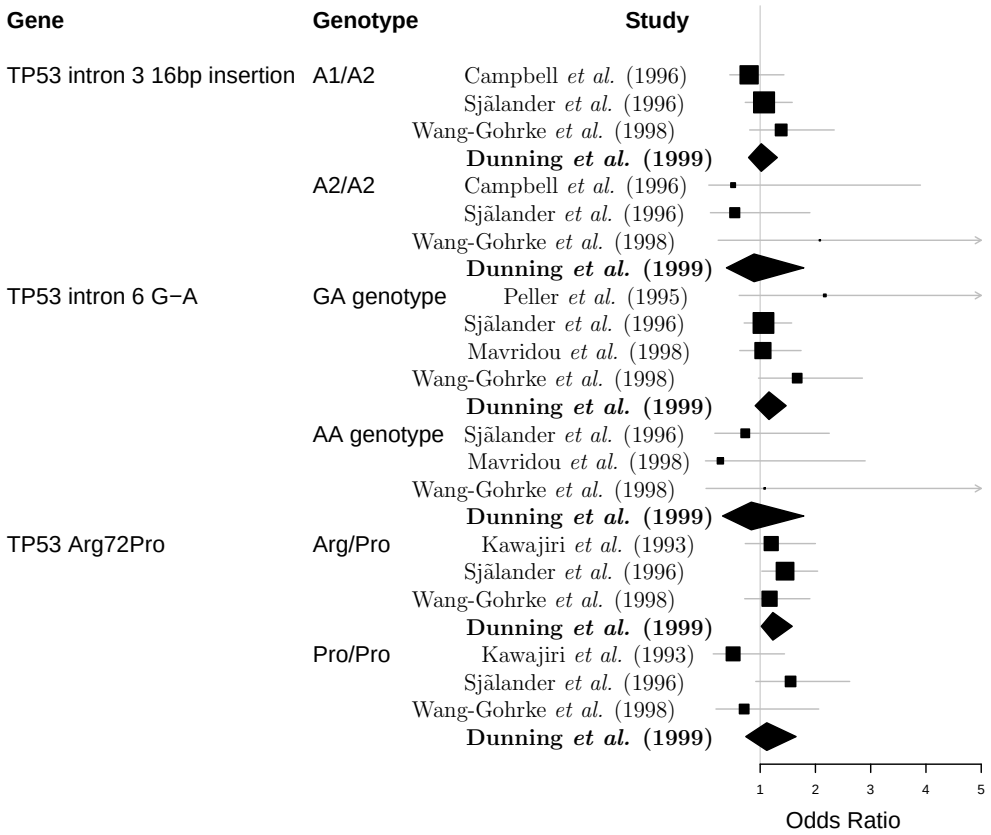


Figure A.3: Forest plot of odds ratio estimates for the gene TP53, with the results of joint analyses by Dunning *et al.* (1999). Horizontal bars indicate 95% confidence intervals.

Appendix B

Intensities of Death and Critical Illness

The intensities $\mu^{OCI}(x)$ and $\mu^{D}(x)$ in Figure 2.3 were taken from Gutiérrez & Macdonald (2003).

For $\mu^{OCI}(x)$, the authors sourced a variety of medical and demographic statistics and fit non-linear functions to intensities of cancer, heart attack and stroke for males and females. The cancer rates were based on the registrations between 1990 and 1992 and were obtained from ONS (1999). Since a typical CI policy requires 28 days of survival after CI onset in order to claim, the rates of heart attack and stroke were adjusted accordingly. For cancer, survival past 28 days from diagnosis is common so no adjustment was necessary for this CI. After summing these rates, they increased the total by 15% to account for minor causes of CI insurance claims (consistent with Macdonald, Waters & Wekwete (2003b)). Figure B.1 shows $\mu^{OCI}(x)$ for males and females.

The rate of mortality, $\mu^{D}(x)$, was based on the English Life Tables Number 15 (ELT15). The rates for male and female were reduced by the proportion of deaths caused by diseases resulting in a CI claim, then the 28 day mortality following heart attack and stroke was added back on. Figure B.2 shows $\mu^{D}(x)$ for males and females.

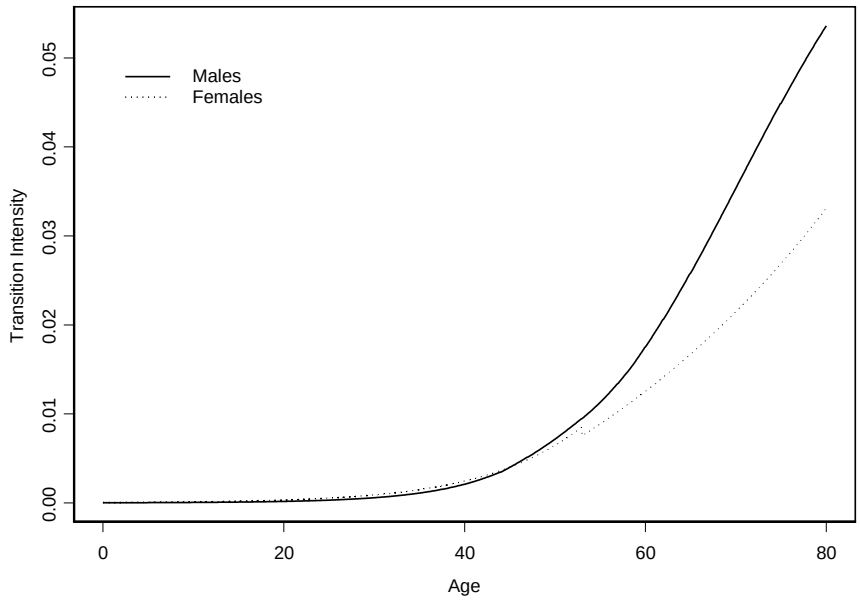


Figure B.1: Incidence rates of other critical illnesses for males and females

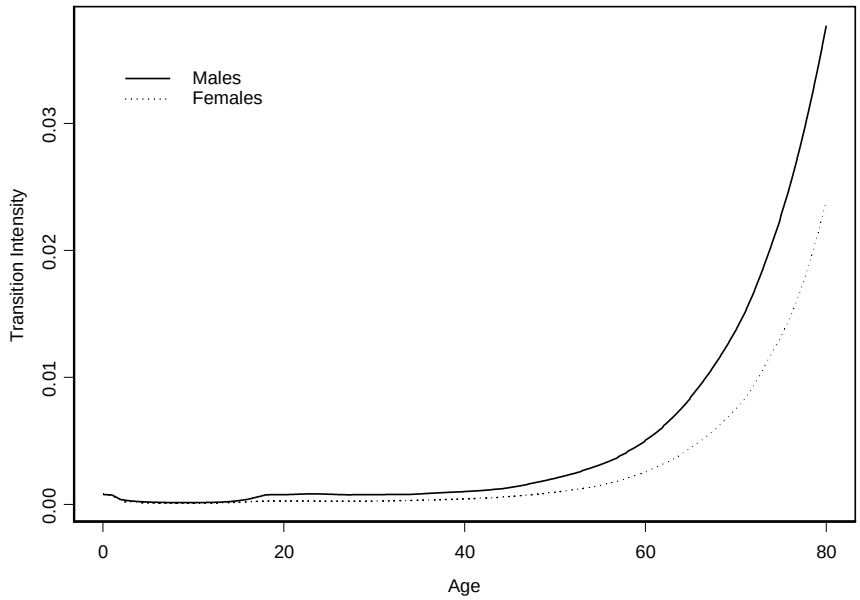


Figure B.2: Mortality rates, based on ELT15, with mortality after CI removed, for males and females

References

Al-Saffar, M. & Foulkes, W.D. (2002). Hereditary ovarian cancer resulting from a non-ovarian cancer cluster region (OCCR) BRCA2 mutation: is the OCCR useful clinically?. *Journal of Medical Genetics*, **39**, 68–70.

Andersen, P.K., Borgan, Ø., Gill, R.D., & Keiding, N. (1993). *Statistical models based on counting processes*. Springer-Verlag, New York.

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Ponder, B.J., & Easton, D.F. (2001). Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology*, **21**, 1–18.

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., & Easton, D.F. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer*, **86**, 76–83.

Antoniou, A.C., Pharoah, P.P.D., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, A., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., Anton-Culver H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tulinius, H., Thorlacius, S., Eerola, H., Nevanlinna, H., Syrjäkoski, K., Kallioniemi, O.-P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D. G., & Easton, D.F. (2003). Average risks of breast and ovarian cancer associated with mutations in BRCA1 or BRCA2 detected in case series unselected for family history: A combined analysis of 22 studies. *American Journal of Human Genetics*, **72**, 1117–1130.

Arking, D.E., Atzmon, G., Arking, A., Barzilai, N., & Dietz, H.C. (2005). Association between a functional variant of the KLOTHO gene and high-density lipoprotein cholesterol, blood pressure, stroke, and longevity. *Circulation Research*, **96**, 412–418.

Australian Institute of Health and Welfare (1999). Breast cancer in Australian women 1982–1996. *Australian Institute of Health and Welfare, Canberra*.

Bailey, L.R., Roodi, N., Verrier, C.S., Yee, C.J., Dupont, W.D. & Parl, F.F. (1998). Breast cancer and CYP1A1, GSTM1, and GSTT1 polymorphisms: evidence of a lack of association in Caucasians and African-Americans. *Cancer Research*, **58**, 65–70.

Bradbury, J. (2002). Could polygenic analysis improve breast-cancer screening?. *The Lancet*, **359:9309**, 857.

Buchert, E.T., Woosley, R.L., Swain, S.M., Oliver, S.J., Coughlin, S.S., Pickle, L., Trock, B. & Riegel, A.T. (1993). Relationship of CYP2D6 (debrisoquine hydroxylase) genotype to breast cancer susceptibility. *Pharmacogenetics*, **3**, 322–327.

Burden, R.L. & Faires, J.D. (1997). *Numerical analysis*. Sixth Edition, Brooks/ Cole.

Burton, P.R., Palmer, L.J., Jacobs, K., Keen, K.J., Olson, J.M. & Elston, R.C. (2001). Ascertainment adjustment: Where does it take us?. *American Journal of Human Genetics*, **67**, 1505–1514.

Cairns, A. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**, 313–330.

Campbell, I.G., Eccles, D.M., Dunn, B., Davis, M. & Leake, V. (1996). p53 polymorphism in ovarian and breast cancer. *Lancet*, **347**, 393–394.

Cannings, C. & Thompson, E.A. (1977). Ascertainment in the sequential sampling of pedigrees. *Clinical Genetics*, **12**, 208–212.

Cannings, C., Thompson, E.A., & Skolnick, M.H. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability*, **10:1**, 26–61.

Charrier, J., Maugard, C.M., Le Mevel, B. & Bignon, Y.J. (1999). Allelotype influence at glutathione *S*-transferase M1 locus on breast cancer susceptibility. *British Journal of Cancer*, **79**, 346–353.

Cui, J., Antoniou, A.C., Dite, G.S., Southey, M.S., Venter, D.J., Easton, D.F., Giles, G.G., McCredie, M.R.E. & Hopper, J.L. (2001). After BRCA1 and BRCA2—what next? Multifactorial segregation analysis of three-generation, population-based Australian families affected by female breast cancer. *American Journal of Human Genetics*, **68**, 420–431.

De Benedictis, G., Tan, Q., Jeune, B., Christensen, K., Ukraintseva, S.V., Bonafè, M., Franceschi, C., Vaupel, J.W., & Yashin, A.I. (2001). Recent advances in human gene-longevity association studies. *Mechanisms of Ageing and Development*, **122**, 909–920.

Doherty, N.A. & Thistle, P.D. (1996). Adverse selection with endogenous information in insurance markets. *Journal of Public Economics*, **63**, 83–102.

Dunning, A., Healey, C.S., Pharoah, P.D.P., Foster, N., Easton, D.F., Day, N.E. & Ponder, B.A.J. (1998). No association between a polymorphism in the steroid metabolism gene CYP17 and risk of breast cancer. *British Journal of Cancer*, **77**, 2045–2047.

Dunning, A., Healey, C.S., Pharoah, P.D.P., Teare, M.D., Ponder, B.A. J. & Easton, D.F. (1999). A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, **8**, 843–854.

Easton, D.F. (1999). How many more breast cancer predisposition genes are there?. *Breast Cancer Research*, **1**, 14–17.

Easton, D.F. (2005). Finding new breast cancer genes. *Presentation at University of Sheffield*.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Stuewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., the SEARCH collaborators, Meyer, K.B., Haiman, C.A., Kolonel, L.K.,

Henderson, B.E., Marchand, L.L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C-Y., Wu, P-E., Wang, H-C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closa, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K-Y., Noh, D-Y., Ahn, S-H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y-L., Bogdanova, N., Schürmann, P., Dörk, T., Tollenaar, R.A.E.M., Jacobi, C.E., Devilee, P., Klijn, J.G.M., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W.R., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y-D., Spurdle, A.B., Beesley, J., Chen, X., kConFab, AOCS Management Group, Mannermaa, A., Kosma, V-M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R. & Ponder, B.A.J. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447 (7148)**, 1087–1093.

Eccles, D.M., Evans, D.G.R., & Mackay, J. (2000). Guidelines for a genetic risk based approach to advising women with a family history of breast cancer. *Journal of Medical Genetics*, **37**, 203–209.

Eisenhauer, J.G. & Ventura, L. (2003). Survey measures of risk aversion and prudence. *Applied Economics*, **35:13**, 1477–1484.

Elston, R.C. (1973). Ascertainment and age of onset in pedigree analysis. *Human Heredity*, **23**, 105–112.

Falconer, D.S. (1981). *Introduction to quantitative genetics*. Edition 2, Longman, New York.

Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., van Duijn, C.M., & APOE and Alzheimer's Disease Meta Analysis Consortium (1997). Effects of age, gender and ethnicity on the association between apolipoprotein E genotype and Alzheimer's disease. *Journal of the American Medical Association*, **278**, 1349–1356.

Feigelson, H.S., Coetzee, G.A., Kolonel, L.N., Ross, R.K. & Henderson, B.E. (1997). A polymorphism in the CYP17 gene increases the risk of breast cancer. *Cancer Research*, **57**, 1063–1065.

Finkel, T., Serrano, M. & Blasco, M.A. (2007). The common biology of cancer and ageing. *Nature*, **448**, 767–773.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A.J., Gayther, S.A., Birch, J.M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, A., Hamann, U., Haites, N., Scott, R.J., Maugard, C.M., Vasen, H., Seitz, S., Cannon-Albright, L.A., Schofield, A., Zelada-Hedman, M., and the Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *American Journal of Human Genetics*, **62**, 676–689.

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

Gui, E.H. & Macdonald, A.S. (2002b). Early-onset Alzheimer's disease, critical illness insurance and life insurance. *Genetics and Insurance Research Centre Research Report, Heriot-Watt University*, **2(2)**, 1–31.

Gui, E.H., Lu, B., Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2006). The genetics of breast and ovarian cancer III: A new model of family

history with insurance applications. *Scandinavian Actuarial Journal*, **2006**, 338–367.

Gutiérrez, M.C. & Macdonald, A.S. (2002a). Huntington's disease and insurance I: A model of Huntington's disease. *Genetics and Insurance Research Centre Research Report, Heriot-Watt University*, **2(3)**, 1–28.

Gutiérrez, M.C. & Macdonald, A.S. (2002b). Huntington's disease and insurance II: Critical illness and life insurance. *Genetics and Insurance Research Centre Research Report, Heriot-Watt University*, **2(4)**, 1–33.

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

Gutiérrez, M.C. & Macdonald, A.S. (2004). Huntington's disease, critical illness insurance and life insurance. *Scandinavian Actuarial Journal*, **4**, 279–313.

Gutiérrez, M.C. & Macdonald A.S. (2007). Adult polycystic kidney disease and insurance: A case study in genetic heterogeneity. *North American Actuarial Journal*, **11:1**, 90–118.

Haiman, C.A., Hankinson, S.E., Speizer, F.E. & Hunter, D.J. (1999). A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk. *Proceedings of the American Association for Cancer Research*, **40**, 194.

Hanley, J., A. (2001). A heuristic approach to the formulas for population attributable fraction. *Journal of Epidemiology and Community Health*, **55**, 508–514.

Harries, L.W., Stubbins, M.J., Forman, D., Howard, G.C.W. & Wolf, C.R. (1997). Identification of genetic polymorphisms at the glutathione *S*-transferase Pi locus and association with susceptibility to bladder, testicular and prostate cancer. *Carcinogenesis*, **18**, 641–644.

Hayden, K.M., Zandi, P.P., Lyketsos, C.G., Tschanz, J.T., Norton, M.C., Khachaturian, A.S., Pieper, C.F., Welsh-Bohmer, K.A., & Breitner, J.C.S. (2005). Apolipoprotein E genotype and mortality: Findings from the Cache County study. *Journal of the American Geriatrics Society*, **53**, 935–942.

Healey, C.S., Dunning, A.M., Durocher, F., Teare, D., Easton, D.F. & Ponder, B.A.J. (1999). Polymorphisms in the human aromatase cytochrome P450 gene (CYP19) and breast cancer risk. *Carcinogenesis*, **21:2**, 189–193.

Helzlsouer, K.J., Huang, H-Y., Strickland, P.T., Hoffman, S., Alberg, A.J., Comstock, G.W. & Bell, D.A. (1998a). Association between CYP17 polymorphism and the development of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **7**, 945–949.

Helzlsouer, K.J., Selmin, O., Huang, H-Y., Strickland, P.T., Hoffman, S., Alberg, A.J., Watson, M., Comstock, G.W. & Bell, D. (1998b). Association between glutathione *S*-transferase M1, P1, and T1 genetic polymorphisms and development of breast cancer. *Journal of the National Cancer Institute*, **90**, 512–518.

Herskind, A.M., McGue, M., Holm, N.V., Sorensen, T.I., Harvald, B., & Vaupel, J.W. (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs 1870–1900. *Human Genetics*, **97:3**, 319–323.

Hodge, S.E. & Vieland, V.J. (1996). The essence of single ascertainment. *Genetics*, **144**, 1215–1223.

Hoem, J.M. (1988). The versatility of the Markov chain as a tool in the mathematics of life insurance. *Transactions of the 23rd International Congress of Actuaries, Helsinki* **S**, 171–202.

Hoy, M. & Witt, J. (2005), Welfare effects of banning genetic information in the life insurance market: The case of BRCA1/2 genes, *University of Guelph Discussion Paper*, **2005-5**.

Human Genetics Commission (2000). *Whose hands on your genes?*. www.hgc.gov .uk.

Huober, J., Bertram, B., Petru, E., Kaufmann, M. & Schmahl, D. (1991). Metabolism of debrisoquine and susceptibility to breast cancer. *Breast Cancer Research and Treatment*, **18**, 43–48.

Kawajiri, K., Nakachi, K., Imai, K., Watanabe, J. & Hayashi, S. (1993). Germ line polymorphisms of p53 and CYP1A1 genes involved in human lung cancer. *Carcinogenesis*, **14**, 1085–1089.

Kristensen, V.N., Andersen, T.I., Lindblom, A., Erikstein, B., Magnus, P. & Børresen-Dale, A.L. (1998). A rare CYP19 (aromatase) variant may increase the risk of breast cancer. *Pharmacogenetics*, **8**, 43–48.

Ladero, J.M., Benitez, J., Jara, C., Llerena, A., Valdivielso, M.J., Munoz, J.J. & Vargas, E. (1991). Polymorphic oxidation of debrisoquine in women with breast cancer. *Oncology*, **48**, 107–110.

Ladona, M.G., Abildua, R.E., Ladero, J.M., Roman, J.M., Plaza, M.A., Agundez, J.A., Munoz, J.J. & Benitez, J. (1996). CYP2D6 genotypes in Spanish women with breast cancer. *Cancer Letters*, **99**, 23–28.

Lange, K. (1997). An approximate model of polygenic inheritance. *Genetics*, **147**, 1423–1430.

Lavigne, J.A., Helzlsouer, K.J., Huang, H-Y., Strickland, P.T., Bell, D.A., Selmin, O., Watson, M.A., Hoffman, S., Comstock, G.W. & Yager, J.D. (1997). An association between the allele coding for a low activity variant of catechol-*O*-methyltransferase and the risk of breast cancer. *Cancer Research*, **57**, 5493–5497.

Le Grys, D. (1997). Actuarial considerations on genetic testing. *Philosophical Transactions of the Royal Society B*, **352**, 1057–1061.

Lemaire, J., Subramanian, K., Armstrong, K., & Asch, D.A. (2000). Pricing term insurance in the presence of a family history of breast cancer. *North American Actuarial Journal*, **4**, 75–87.

Levin, M.L. (1953). The occurrence of lung cancer in man. *Acta Unio Internationalis Contra Cancrum*, **9**, 531–541.

Lu, L., Macdonald, A.S., & Waters, H.R. (2007). Premium rates based on genetic studies: How reliable are they?. *To appear in Insurance: Mathematics and Economics*.

Macdonald, A.S. (1997). How will improved forecasts of individual lifetimes affect underwriting?. *Philosophical Transactions of the Royal Society*, **352**, 1067–1075.

Macdonald, A.S. (1999). Modeling the impact of genetics on insurance. *North American Actuarial Journal*, **3:1**, 83–101.

Macdonald, A.S. (2000). *Human genetics and insurance issues*. In Bio-ethics for the New Millennium, edited by I. Torrance. St. Andrew Press.

Macdonald, A.S. & Pritchard, D.J. (2000). A mathematical model of Alzheimer's disease and the ApoE gene. *ASTIN Bulletin*, **30**, 69–110.

Macdonald, A.S. & Pritchard, D.J. (2001). Genetics, Alzheimer's disease, and long-term care insurance. *North American Actuarial Journal*, **5:2**, 54–78.

Macdonald, A.S., Pritchard, D.J. & Tapadar, P. (2006). The impact of multifactorial genetic disorders on critical illness insurance: A simulation study based on UK Biobank. *ASTIN Bulletin*, **36**, 311–346.

Macdonald, A.S. & Tapadar, P. (2006). Multifactorial genetic disorders and adverse selection: Epidemiology meets economics. *Genetics and Insurance Research Centre Research Report, Heriot-Watt University*, **6(6)**, 1–27.

Macdonald, A.S., Waters, H.R., & Wekwete, C.T. (2003a). The genetics of breast and ovarian cancer I: A model of family history. *Scandinavian Actuarial Journal*, **1**, 1–27.

Macdonald, A.S., Waters, H.R., & Wekwete, C.T. (2003b). The genetics of breast and ovarian cancer II: A model of critical illness insurance. *Scandinavian Actuarial Journal*, **1**, 28–50.

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2005a). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting I: The model. *North American Actuarial Journal*, **9:1**, 13–40.

Macdonald, A.S., Waters, H.R. & Wekwete, C.T. (2005b). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting II: Applications. *North American Actuarial Journal*, **9:1**, 41–56.

Mavridou, D., Gornall, R., Campbell, I.G. & Eccles, D.M. (1998). TP53 intron 6 polymorphism and the risk of ovarian and breast cancer. *British Journal of Cancer*, **77**, 676–677.

McGue, M., Vaupel, J.W., Holm, N., & Harvald, B. (1993). Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *Journal of Gerontology*, **48:6**, 237–244.

Millikan, R.C., Pittman, G.S., Tse, C.K., Duell, E., Newman, B., Savitz, D., Moorman, P.G., Boissy, R.J. & Bell, D.A. (1998). Catechol-*O*-methyltransferase and breast cancer risk. *Carcinogenisis*, **19**, 1943–1947.

National Breast Cancer Centre (2002). Ovarian cancer in Australian women. *National Ovarian Cancer Centre*.ONS1999Cancer 1971–1997CD-ROMOffice for National Statistics, London

Peller, S., Kopilova, Y., Slutzki, S., Halevy, A., Kvitko, K. & Rotter, V. (1995). A novel polymorphism in intron 6 of the human p53 gene: a possible association with cancer predisposition in malignant and benign breast disease. *European Journal of Cancer*, **26**, 790–792.

Pontin, J.E., Hamed, H., Fentiman, I.S. & Idle, J.R. (1998). Cytochrome p450dbl phenotypes in malignant and benign breast disease. *European Journal of Cancer*, **26**, 790–792.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (2002). *Numerical recipes in C++. The art of scientific computing.* Second Edition, Cambridge University Press.

Rabinowitz, D. (1996). A pseudolikelihood approach to correcting for ascertainment bias in family studies. *American Journal of Human Genetics*, **59**, 726–730.

Rebbeck, T.R. (1999). Inherited genetic predisposition in breast cancer. A population-based perspective. *Cancer*, **86**, 2493–2501.

Rennert, G., Bisland-Naggan, S., Barnett-Griness, O., Bar-Joseph, N., Zhang, S., Rennert, H.S. & Narod, S.A. (2007). Clinical outcomes of

breast cancer in carriers of BRCA1 and BRCA2 mutations. *New England Journal of Medicine*, **357(2)**, 115–123.

Ropka, M.E., Wenzel, J., Phillips, E.K., Siadaty, M. & Philbrick, J.T. (2006). Uptake rates for breast cancer genetic testing: A systematic review. *Cancer Epidemiology Biomarkers and Prevention*, **15:5**, 840–855.

Santoro, A., Salvioli, S., Raule, N., Capri, M., Sevini, F., Valensin, S., Monti, D., Bellizzi, D., Passarino, G., Rose, G., De Benedictis, G. & Franceschi, C. (2006). Mitochondrial DNA involvement in human longevity. *Biochimica et Biophysica Acta*, **1757:9–10**, 1388–1399.

Select Committee on Economic Affairs (2004). *Aspects of the economics of an ageing population*. House of Lords, Session 2002–03, 4th Report.

Seigelmann-Danieli, N. & Buetow, K.H. (1999). Constitutional genetic variation at the human aromatase gene (Cyp19) and breast cancer risk. *British Journal of Cancer*, **79**, 456–463.

Sjãlnader, A., Birgander, R., Hallmans, G., Cajander, S., Lenner, P., Athlin, L., Beckman, G. & Beckman, L. (1996). p53 polymorphisms and haplotypes in breast cancer. *Carcinogenesis*, **17**, 1313–1316.

Smith, C.A., Moss, J.E., Gough, A.C., Spurr, N.K. & Wolf, C.R. (1992). Molecular genetic analysis of the cytochrome P450-debrisoquine hydroxylase locus and association with cancer susceptibility. *Environmental Health Perspectives*, **98**, 107–112.

Strachan, T. & Read, A.P. (2004). *Human Molecular Genetics 3*. Garland Publishing.

Struewing, J.P. (2004). Genomic approaches to identifying breast cancer susceptibility factors. *Breast Disease*, **19**, 3–9.

Subramanian, K., Lemaire, J., Hershey, J.C., Pauly, M.V., Armstrong, K., & Asch, D.A. (1999). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *The Journal of Risk and Insurance*, **66**, 531–550.

Taioli, E., Trachman, J., Chen, X., Toniolo, P. & Garte, S.J. (1995). A CYP1A1 restriction fragment length polymorphism is associated with breast cancer in African-American women. *Cancer Research*, **55**, 3757–3758.

Tamimi, R. (2006). Single nucleotide polymorphisms and breast cancer: not yet a success story. *Breast Cancer Research*, **8(4)**, 108.

Tan, Q., De Benedictis, G., Yashin, A.I., Bonafè, M., DeLuca, M., Valensin, S., Vaupel, J.W., & Franceschi, C. (2001). Measuring the genetic influence in modulating the human life span: gene-environment interaction and the sex-specific genetic effect. *Biogerontology*, **2**, 141–153.

Thompson, D.J. & Easton, D.F. (2001). Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. *American Journal of Human Genetics*, **68**, 410–419.

Wang-Gohrke, S., Rebbeck, T.R., Besenfelder, W., Kreienberg, R. & Runnebaum, I.B. (1998). p53 germline polymorphisms are associated with an increased risk for breast cancer in German women. *Anticancer Research*, **18**, 2095–2099.

Wekwete, C.T. (2002). Genetics and critical illness insurance underwriting: models for breast cancer and ovarian cancer and for coronary heart disease. *PhD thesis*, Heriot-Watt University.

Weston, A., Pan, C-F., Bleiweiss, I.J., Ksieski, H.B., Roy, N., Maloney, N. & Wolff, M.S. (1998). CYP17 genotype and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, **7**, 941–944.

Zhong, S., Wyllie, A.H., Barnes, D., Wolf, C.R. & Spurr, N.K. (1993). Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer. *Carcinogenesis*, **14**, 1821–1824.