GENETICS AND HEALTH INSURANCE

Angus Macdonald
Department of Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical Sciences
Heriot-Watt University, Edinburgh
www.maxwell.ac.uk
“I am not opposed to people knowing their predisposition to an illness. ... I do oppose insurance companies and others taking this into account when they are assessing premiums, the prospects of getting a mortgage and employment.”

(Dr Ian Gibson MP, Daily Mail, 12 October 2000)
Outline

• Review of genetics and insurance
• Genetics yesterday and tomorrow
• Actuarial modelling (1)
  – Epidemiology of multifactorial disorders
  – UK Biobank
  – A simulation study
• Actuarial modelling (2)
  – Breast cancer as a single-gene disorder
  – Breast cancer as a polymorphic disorder
Genetics (and Insurance) of Yesterday

- Family history of Mendelian disorders – clear genetics
- Family history of common diseases – unclear genetics
- DNA-based genetic tests – mid-1990s to now
- The “genetics and insurance debate”
  - unfair discrimination versus adverse selection
  - genetics = precise prediction?
  - argument from a few models e.g. Huntington’s disease
  - strong media focus
Single-Gene Disorders

Gene $\rightarrow$ Disease
Outcomes – the UK as an Example

• Participants
  – Government (HoC, DoH, DTI)
  – Industry (Association of British Insurers)
  – Commissions (HGAC, HGC)
  – Academia (GIRC)

• Outcomes
  – List of “significant” disorders
  – Genetics and Insurance Committee (GAIC)
Moratoria

- Insurers will not ask someone to take a genetic test
- Insurers will not ask about results from research trials
- Insurers will not ask about existing predictive tests
  - Up to £500,000 of life insurance
  - Up to £300,000 of critical illness insurance
  - But only for approved genetic tests
- Use of family history not restricted (compare Sweden)
The Genetics and Insurance Committee (GAIC)

- Insurers may apply to GAIC to be allowed to use *specific* test results, above the limits in the moratorium.
- GAIC will assess:
  - The technical relevance of the test.
  - The clinical relevance of the test – does it predict outcomes?
  - The actuarial relevance of the test – is it material?
- So far, one application (Huntington’s, life insurance).
- Evidence of impact precedes use in underwriting – a precedent for insurance or a one-off?
Genetics of Tomorrow

- Genetics of common diseases
- Gene-gene, gene-environment interactions
- Whole-genome scans
- Genetic arrays
- Large-scale population studies
- Novel mechanisms (epigenetics, RNA interference)
- Genetic therapy
Multifactorial Disorders

Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
Gene 6

Disease

Smoking

Diet

Affluence
Genetics of Tomorrow

- Genetics of common diseases
- Gene-gene, gene-environment interactions
- Whole-genome scans
- Genetic arrays
- Large-scale population studies
- Novel mechanisms (epigenetics, RNA interference)
- Genetic therapy
UK Biobank

- Recruit 500,000 subjects age 40-69
- DNA samples from all subjects
- Lifestyle/medical details collected
- Follow up for 10 years
- Linkage to health records through personal doctor
- Linkage to cancer registries
- Linkage to death registries
UK Biobank

• UK Biobank only collects data
• Analysis is separate (not yet funded)
• Investigators apply to obtain UK Biobank data
• Most analyses will be case-control studies
UK Biobank

- UK Biobank only collects data
- Analysis is separate (not yet funded)
- Investigators apply to obtain UK Biobank data
- Most analyses will be case-control studies

- “Data from the project will not be accessible to the insurance industry or any other similar body” (UK Biobank draft protocol).
Case-Control Studies

• Method of analysing data retrospectively
• Hypothesis: some factor (e.g. genotype) is associated with some outcome (e.g. disease)
• Step 1: collect cases, \( a+c \) in total:
  – \( a \) have genotype G
  – \( c \) have genotype \( g \)
• Step 2: collect undiseased controls, \( b+d \) in total:
  – \( b \) have genotype G
  – \( d \) have genotype \( g \)
Odds Ratios

- Actuaries will only have access to published case-control studies which will report odds ratios.
- The odds of an event with probability $P$ are $P/(1-P)$.
- The odds ratio of an event with probability $Q$ with respect to an event with probability $P$ is $P(1-Q)/Q(1-P)$.
- In the case-control study, $ad/bc$ is an unbiased estimate of the odds ratio of disease, of genotype $G$ compared with genotype $g$. 
Actuarial Use of Odds Ratios

- Actuarial models are based on intensities or forces (or probabilities obtained from them)
- Given a baseline force or intensity (e.g. risk in general population) all we need are relative risks (e.g. for each genotype)
- If $P$ and $Q$ are small then the odds ratio approximates the relative risk
Simulating UK Biobank

- Model of health insurance contract (critical illness)
- Sub-model of common disease risk (heart attack)
- Model of gene-environment interaction
  - Population frequencies and relative risks
- Simulate UK Biobank recruitment phase
  - Allocate 500,000 persons to age, genotype and environment
- Simulate 500,000 lifetimes over 10 years
- We have UK Biobank!
A Simple Critical Illness Insurance Model

- Healthy
  - Heart Attack
  - Other CI
  - Dead
A Simple Heart Attack Model

Healthy \rightarrow \text{Heart Attack} \xrightarrow{\lambda_s} \text{Dead}

\text{Dead} \rightarrow \text{Dead}
Gene-Environment Interaction Model

- Beneficial genotype $g$ and adverse genotype $G$
- Beneficial environment $e$ and adverse environment $E$
- Four strata $ge$, $Ge$, $gE$, $GE$, for each sex
- 10% of population in each of $G$ and $E$ (independently)
- Table shows relative risks in each stratum, with respect to population heart attack risk

<table>
<thead>
<tr>
<th></th>
<th>$G$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>1.3</td>
<td>0.9</td>
</tr>
<tr>
<td>$e$</td>
<td>1.1</td>
<td>0.7</td>
</tr>
</tbody>
</table>
## Five UK Biobank Life Histories

<table>
<thead>
<tr>
<th>ID</th>
<th>Stratum</th>
<th>Sex</th>
<th>Age</th>
<th>Age at HA</th>
<th>Age at Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ge</td>
<td>M</td>
<td>41.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ge</td>
<td>M</td>
<td>58.74</td>
<td>63.89</td>
<td>63.94</td>
</tr>
<tr>
<td>3</td>
<td>ge</td>
<td>M</td>
<td>52.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ge</td>
<td>M</td>
<td>68.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Ge</td>
<td>F</td>
<td>60.94</td>
<td>62.81</td>
<td></td>
</tr>
</tbody>
</table>
Model Epidemiologist and Model Actuary

• Our model epidemiologist obtains funding to do a case control study, and publishes odds ratios:
  – Each sex
  – Each stratum, with respect to stratum ge
  – 5-year age bands

• Our model actuary “extracts” relative risks from odds ratios and parameterises critical illness model
Results – Extra Premiums w.r.t. Stratum $ge$

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Age</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Term 15</td>
<td>Term 25</td>
</tr>
<tr>
<td>$ge$</td>
<td>45</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>$Ge$</td>
<td>45</td>
<td>21%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>16%</td>
<td>14%</td>
</tr>
<tr>
<td>$GE$</td>
<td>45</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>24%</td>
<td>21%</td>
</tr>
</tbody>
</table>
Reliability – GAIC’s Questions

• What would GAIC want to know?
• Does genotype + environment constitute an effective, predictive risk factor?
• Approach:
  – Fix environmental and genetic risk model
  – Simulate 500,000 life histories
  – Do case-control studies and calculate premiums
  – Repeat 1,000 times
Premium Ratings as a Proportion of Premium in Stratum $ge$
Case Control Study With 10,000 Cases
Case-Control Study With 1,000 Cases
Is Our Model Realistic?

- Only 2 genotypes
- Only 2 environmental factors
- Age and sex only other covariates
- Simple multiplicative interaction
- Epidemiologist hits on “correct” model
- Study population free of “noise” and dependencies
Conclusion: Will Biobank Be Relevant?

- UK Biobank can distinguish risk differentials of the order of +50% or more (as it was designed to do)
- Point estimates of these differentials can be used to find premium ratings
- But the distributions of these ratings may not be reliably distinct, in GAIC terms, unless very large numbers of cases are used.

Insurance in the Multifactorial World

• High-throughput genetic arrays will reveal much about complex genetic influences on biological processes – but this is not the same as disease.

• Understanding biological processes better will help to understand disease – but this is not the same as epidemiology.

• Epidemiology will emerge:
  – But it will not be highly predictive, as for single-gene disorders
  – And if subjected to GAIC-like criteria it might fail “reliability”.

Personalised Medicine – Here At Last?

- **Oncotype DX**: 21 gene screen test
  - Algorithm profiles breast cancer recurrence risk
  - Identifies value of chemotherapy
  - Cost $3,400
- Will be paid for by insurers covering 40% of US market
- Insurers’ costs will increase? (cost of tests)
- Insurers’ costs will decrease? (fewer ineffective treatments)

What Will the Press Think?

• The chain from genetic discovery to reliable underwriting is very long and getting longer:
  – Association of genes with disease
  – Understanding complex mechanisms
  – Gene-environment and other interactions
  – Epidemiological studies
  – Moratoria and GAIC-type approval processes

• But the press will not understand this.
• THIS is the actuarial research message.