The Effect of Model Uncertainty on the Pricing of Critical Illness Insurance

George Streftaris

Work with
- Erengul Dodd (Ozkok)
- Howard Waters
- Andrew Stott
- David Wilkie

Funded by the IFoA

Outline:

1. Critical Illness Insurance (CII)
2. Data & Problems
3. Claim delay distribution modelling
4. CI diagnosis rates under uncertainty
5. CI pricing under uncertainty
Critical Illness: Policy description

- Fixed term policy, usually ceasing at age 65
- A fixed sum insured payable on the diagnosis of one of a specified list of illnesses
- Benefit type:
  - Full acceleration (FA): Death is included as a critical illness (88%)
  - Stand alone (SA): Death is not included as a critical illness (12%)
- Covers:
  - Cancer; Death; Heart attack; Stroke; Multiple Sclerosis; Total & permanent disability; Coronary artery bypass graft; Kidney failure; Major organ transplant; Other.

Data

CI data for 1999 – 2005 supplied by CMI:

- Details of policies inforce at the start and end of each year
  - 18,000,000 policy-years of exposure
- Details of claims settled in 1999 – 2005
  - 19,000 claims

Covariates in the data:

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Number of levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Numerical</td>
</tr>
<tr>
<td>Sex</td>
<td>2 (Female = 0)</td>
</tr>
<tr>
<td>Smoker status</td>
<td>2 (NS = 0)</td>
</tr>
<tr>
<td>Policy duration</td>
<td>13</td>
</tr>
<tr>
<td>Office</td>
<td>2 (FA = 0 &amp; SA)</td>
</tr>
<tr>
<td>Benefit type</td>
<td>Numerical</td>
</tr>
<tr>
<td>Benefit amount</td>
<td>2 (Single/ Joint life = 0)</td>
</tr>
<tr>
<td>Policy type</td>
<td>Numerical</td>
</tr>
<tr>
<td>Settlement year</td>
<td>10</td>
</tr>
<tr>
<td>Cause</td>
<td></td>
</tr>
</tbody>
</table>

Diseases covered

Critical illnesses and percentage of claims in 1999 – 2005:

<table>
<thead>
<tr>
<th>Critical Illness</th>
<th>% claims</th>
<th>Critical Illness</th>
<th>% claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>49.0</td>
<td>Total &amp; permanent disability (TPD)</td>
<td>2.6</td>
</tr>
<tr>
<td>Death</td>
<td>17.6</td>
<td>Coronary artery bypass graft (CABG)</td>
<td>2.1</td>
</tr>
<tr>
<td>Heart attack (HA)</td>
<td>11.6</td>
<td>Kidney failure (KF)</td>
<td>0.6</td>
</tr>
<tr>
<td>Stroke</td>
<td>5.4</td>
<td>Major organ transplant (MOT)</td>
<td>0.2</td>
</tr>
<tr>
<td>Multiple sclerosis (MS)</td>
<td>4.3</td>
<td>Other causes</td>
<td>6.6</td>
</tr>
<tr>
<td>Males</td>
<td>57.3</td>
<td>Non-smokers</td>
<td>73.9</td>
</tr>
<tr>
<td>Females</td>
<td>42.7</td>
<td>Smokers</td>
<td>26.1</td>
</tr>
</tbody>
</table>

Source: Continuous Mortality Investigation, UK
Modelling & estimation

- Estimation & smoothing of CI diagnosis rates
  - how do these depend on risk factors?
- Diagnosis is the insured event and there is a delay between diagnosis and settlement
  - diagnosis date often not recorded (18%); need to model it
  - does delay also depend on risk factors?
- The exposure corresponds to claims settled, not to claims diagnosed; need to adjust it
- Premium pricing
- Also take into account uncertainty

Diagnosis to settlement delay

Mean Delay = 185 days; SD Delay = 263 days

Fit a claim delay distribution (CDD):

\[ F(t; x; s; \theta) = \Pr[\text{claim diagnosed at age } x, \text{ cause } s, \text{ covariates } \theta, \text{ will be settled within } t \text{ years}] \]

- Estimate missing dates of diagnosis as:
  - Date of settlement = median of appropriate CDD
  (posterior distribution available)
- Use the CDD to adjust the exposure
- Office-specific growth weights introduced

Observed delay & null model fit
Claim Delay (D) Distribution

Include risk factors in GLM setting:

\[ M_1: \quad D_i \sim LN(\mu_i, \sigma^2) \]

where \[ \mu_i = \delta_0 + \sum_{k=1}^{s} \delta_{ik} = \delta_0 + \delta_k + \delta_{i,k} \]

\[ M_2: \quad D_i \sim \text{Transformed Gamma}(\alpha, \tau, \kappa) \]

\[ f_2(u) = \frac{\kappa^{\alpha}(\kappa u)^{\alpha-1} \exp(-\kappa u)}{\alpha^{\alpha} \Gamma(\alpha)} \quad , \quad E(D_i) = \exp(\kappa) \]

where \( \kappa \) as above and \( \alpha \) given as function of \( \eta, \alpha, \tau \).

Claim Delay (D) Distn (cont)

\[ M_3: \quad D_i \sim \text{Burr}(\alpha, \tau, \kappa) \]

\[ f_3(u) = \frac{\kappa^{\alpha}(\kappa u)^{\alpha-1} \exp(-\kappa u)}{\alpha^{\alpha} \Gamma(\alpha) u^{\alpha+1}} \quad , \quad E(D_i) = \exp(\kappa) \]

with \( \kappa \) given as function of \( \eta, \alpha, \tau \).

\[ M_4: \quad D_i \sim \text{Generalized beta}(\alpha, \tau, \gamma, \kappa) \]

\[ f_4(u) = \frac{\Gamma(1+\gamma)}{\Gamma(\alpha+\gamma)} \frac{\kappa^{\alpha+\gamma}(\kappa u)^{\alpha+\gamma-1} \exp(-\kappa u)}{\alpha^{\alpha+\gamma} (1+(\kappa u)^{\alpha+\gamma})^{\alpha+\gamma+1}} \quad , \quad E(D_i) = \exp(\kappa) \]

again, with \( \kappa \) given as function of \( \eta, \alpha, \tau, \gamma \).

Claim Delay (D) Distn (cont)

Fit the 4 models under a Bayesian framework using MCMC:

- Compare fit using Deviance Information Criterion:

<table>
<thead>
<tr>
<th>Model</th>
<th>LN</th>
<th>Tr Gamma</th>
<th>Burr</th>
<th>Gen Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIC</td>
<td>194,336</td>
<td>193,023</td>
<td>191,262</td>
<td>190,992</td>
</tr>
</tbody>
</table>

- Model fit also compared using posterior predictive checking.
Risk factor estimates

Illness coefficient estimates

Prediction of Claim Delay

Find ‘best’ predictive model using Bayesian variable selection – Ozkok et al. (2012a))

<table>
<thead>
<tr>
<th>Covariate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefit type</td>
<td>FA</td>
<td>FA</td>
<td>FA</td>
<td>FA</td>
<td>FA</td>
</tr>
<tr>
<td>Policy type</td>
<td>JL</td>
<td>JL</td>
<td>JL</td>
<td>JL</td>
<td>JL</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>50k</td>
<td>50k</td>
<td>50k</td>
<td>50k</td>
<td>50k</td>
</tr>
<tr>
<td>Office</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Cause</td>
<td>Cancer</td>
<td>Cancer</td>
<td>Cancer</td>
<td>Death</td>
<td>FPD</td>
</tr>
<tr>
<td>$E[</td>
<td>d</td>
<td>]$ (days)</td>
<td>174 (167, 182)</td>
<td>156 (146, 166)</td>
<td>195 (187, 204)</td>
</tr>
</tbody>
</table>

Scenario 1: reference;  Red: changes from reference
Estimation of CI diagnosis rates

- Non-recorded diagnosis dates estimated through CDD model
- Suppose Office 1 contributes data for 2000 to 2003. For this office, let:

\[ \theta \]

be a set of covariates, including office

\[ \lambda^{(i)}_x \]

be the diagnosis inception rate for cause \( i \) at age \( x \) with covariates

\[ \psi \]

be the number of policies (age \( x \), covariates \( \theta \)) insurable at time \( u \), \( 0 \leq u \leq 4 \)

\[ N^{(i)}(x; \theta) \]

be the number of claims (cause \( i \), age \( x \), covariates \( \theta \)) diagnosed and settled in 2000 – 2003

\[ N^{(i)}(x; \theta) \sim \text{Poisson} \left( \lambda^{(i)}_x \int_{u=0}^{4} \psi(u; x; \theta) e^{\psi(u; x; \theta) du} \right) \]

Model

For all causes:

- Healthy
- Insured event

Have also considered:

- Healthy
- Diagnosed with a CI
- Dead

Rate smoothing (graduation)

Gompertz-Makeham-Cox-type Model:

\[ \lambda^{(i)}_{x} = \lambda^{(i)}_{1}(x) + \exp(\lambda^{(i)}_{2}(x)) \exp(\beta z^T) \]

where \( \lambda^{(i)}_{1}(x) \) is a polynomial function of age only; \( i = 1, 2 \)

\[ \lambda^{(i)}_{1}(x) \equiv 0 \text{ for each cause except death} \]

\[ \lambda^{(i)}_{2}(x) = 0 \rightarrow \text{log-linear (Cox-type) model} \text{ for } \lambda^{(i)}_{x} \]
Fit model – perform variable selection using BIC

<table>
<thead>
<tr>
<th>Cause</th>
<th>Predictive model (covariates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGB</td>
<td>Age  Sex  Smoker</td>
</tr>
<tr>
<td>Cancer</td>
<td>Age  Sex  Year  Smoker Age=5m</td>
</tr>
<tr>
<td>Death</td>
<td>Age  Sex  Smoker Age=5m</td>
</tr>
<tr>
<td>Heart Attack</td>
<td>Age  Sex  Smoker Age=5m</td>
</tr>
<tr>
<td>Kidney Fail</td>
<td>Age  Sex  Smoker Age=5m</td>
</tr>
<tr>
<td>MS</td>
<td>Sex  Smoker Pol Durn</td>
</tr>
<tr>
<td>Other</td>
<td>Age  Sex  Office  Benefit type</td>
</tr>
<tr>
<td>Stroke</td>
<td>Age  Year  Smoker Age=5m</td>
</tr>
<tr>
<td>TPD</td>
<td>Age  Year  Smoker Pol Durn</td>
</tr>
</tbody>
</table>

Claim rates for Non-smokers, Pol Durn 0, All causes: **Burr v LN**

Claim rates for **Smokers**, Pol Durn 0, All causes: **Burr v LN**

Comparison of diagnosis rates

Comparison of diagnosis rates (2)
**Sensitivity to diagnosis estimates**

Relative rates (divided by rate obtained with Median of CDD)
- Rate using 97.5th percentile of Burr or LN CDD
- CIs derived using parametric bootstrap (Burr v LN)

![Graph showing sensitivity to diagnosis estimates]

**Premium pricing - derivation**

Annual premium, paid at constant rate, n-year term
- Can be calculated using
  \[\text{Net Premium} = \text{Benefit Amount} \times \frac{\int_0^n v^s d\nu_s \lambda_{x+i} dt}{\int_0^n v^s d\nu_s dt}\]
  where
  - \(\nu_s\): survival probability
  - \(\lambda_{x+i}\): total claim rate at age \(x+i\)
  - \(v\): discount factor
- Then bootstrap distribution of \(\lambda_s\) used to derive CIs for premiums

**Premium pricing - comparisons**

Age 40, Pol Durn 0, All causes, Benefit amount £100k, \(i = 3\%\)

<table>
<thead>
<tr>
<th>LN CDD</th>
<th>Burr CDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-smokers</td>
<td>Non-smokers</td>
</tr>
<tr>
<td><strong>Term</strong></td>
<td><strong>Net premium rate</strong></td>
</tr>
<tr>
<td>5-years</td>
<td>156.05</td>
</tr>
<tr>
<td>25-years</td>
<td>373.69</td>
</tr>
<tr>
<td>Smokers</td>
<td>Smokers</td>
</tr>
<tr>
<td>5-years</td>
<td>239.70</td>
</tr>
<tr>
<td>25-years</td>
<td>714.03</td>
</tr>
</tbody>
</table>
Premium pricing - comparisons

Age 40, Pol Durn 0, All causes, Benefit amount £100k, \( i = 3\% \)

\textit{(Burr v LN CDD)}

Non-smokers  
Smokers

Premiums with cause-specific model

Have also considered model for specific causes  
\textit{(this can also distinguish between FA and SA policies)}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{cause-specific-model.png}
\end{figure}

Useful for comparing rates or premiums for certain causes; or when a specific illness is excluded – e.g. TPD

Premiums with TPD not covered

Age 40, Pol Durn 0, Benefit amount £100k, \( i = 3\% \)

\textit{(Burr v LN CDD)}

Non-smokers  
Smokers
Summary

- Delay between diagnosis and settlement in CII is important (e.g. IBNR, IBNS)
- Have developed delay model: depends on risk factors
- Bayesian analysis accounts for non-recorded diagnosis dates
- 4-parameter G.Beta distn fits data best – followed by 3-parameter Burr
- CII rates and premiums estimated & smoothed
  - including parameter and model uncertainty
- Estimates of CDD are model-sensitive
- But claim rates and premiums are not

More details in:


