Predictive modelling in action: 
(Life) lessons from around the world
Adèle Groyer & Alastair Gerrard
In the fast lane

• 80% of surveyed US personal automobile insurers currently use predictive modelling in underwriting, risk selection, rating or pricing.

• 45% of personal auto carriers use or will use usage-based insurance

Towers Watson 2013 Predictive Modeling Benchmarking Survey
How far down the road are life insurers?

- US Predictive Modeling Industry Survey 2013
- International Predictive Modelling Survey 2014
Recap: what is predictive modelling?

“Predictive modeling can be defined as the analysis of large data sets to make inferences or identify meaningful relationships, and the use of these relationships to better predict future events. It uses statistical tools to separate systematic patterns from random noise, and turns this information into business rules, which should lead to better decision making.”

Topics covered by the surveys

• Current and planned use of predictive modelling
• Specific applications in use or with greatest potential use
• Data availability
Current and planned use of predictive modelling
Andean Countries: Colombia, Ecuador, Peru
Caribbean: Puerto Rico, Dominican Republic
Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Panama
Southern Cone: Paraguay, Chile
Current and planned use of models

- Not currently using predictive modelling, future plans not disclosed
- Unlikely to use predictive modelling
- Unlikely to use predictive modelling imminently but may do so within the next 5 years
- Likely to use predictive modelling within the next couple of years
- Already using predictive modelling

U.S. results as at 2013, others at 2014
Distribution channel

U.S. responses are all individual insurance but no further details are available. Categorised as unknown channel.
Current model use by channel

- Direct to customer
- Own sales force
- Independent brokers
- Via banks
- Group schemes

Mentioned in single or multi-channel
Single channel only

12 February 2015
Who developed the model

All countries except United States:
- Internally created model only: 7%
- Purchased modelling service only: 7%
- Both internally created and purchased models: 87%

United States only:
- Internally created model only: 50%
- Purchased modelling service only: 50%

12 February 2015
Model performance

All regions

- As expected or better: 14
- Too soon to tell: 1
- Worse than expected: 20

12 February 2015
### Reasons for not using predictive models

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of mentions (excl. U.S.) N = 81</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low business volumes / lack of data</td>
<td>10</td>
</tr>
<tr>
<td>Low priority / cost high vs benefit</td>
<td>6</td>
</tr>
<tr>
<td>Lack of expertise</td>
<td>4</td>
</tr>
<tr>
<td>Lack of resources</td>
<td>3</td>
</tr>
<tr>
<td>Data issues in group business</td>
<td>2</td>
</tr>
<tr>
<td>Data privacy concerns</td>
<td>1</td>
</tr>
<tr>
<td>Not relevant to type of business</td>
<td>1</td>
</tr>
</tbody>
</table>

Among U.S. insurers who had not implemented predictive models
- 89% were concerned that there was not enough proof of accuracy
- 33% said it was too expensive
Interest in specific applications
Model applications in use

- Identify prospects more likely to lapse
- Identify prospects more likely to buy
- Target marketing efforts
- Identify new risk or rating factors
- Distribution management
- Inform product differentiation
- Develop targeted underwriting requirements
- Make other pricing refinements
- Reduce / simplify underwriting
- Identify fraud / misrepresentation
- Triage claims
- Refine underwriting decisions
- Do away with underwriting
- Other claims management

Number of mentions

Retention, marketing and distribution
Underwriting and claims

12 February 2015
Question not included in U.S. survey
Model application in use by country

- Identify prospects more likely to lapse
- Identify prospects more likely to buy
- Target marketing efforts
- Distribution management
- Identify new risk or rating factors
- Inform product differentiation
- Make other pricing refinements
- Targeted underwriting requirements
- Triage claims
- Identify fraud / misrepresentation
- Reduce / simplify underwriting
- Refine underwriting decisions
- Other claims management
- Do away with underwriting

**COUNTRIES**
- Australia
- New Zealand
- South Africa / Rest of Africa
- UK / Ireland
- Latin America - Andean Countries
- Latin America - Brazil
- Latin America - Caribbean
- Latin America - Central America
- Latin America - Mexico
- Latin America - Southern Cone

12 February 2015
Model applications with greatest potential (excl. United States)

- Reduce / simplify underwriting
- Identify new risk or rating factors
- Identify prospects more likely to lapse
- Identify prospects more likely to buy
- Target marketing efforts
- Targeted underwriting requirements
- Inform product differentiation
- Make other pricing refinements
- Refine underwriting decisions
- Identify fraud / misrepresentation
- Distribution management
- Triage claims
- Do away with underwriting
- Other claims management

Number of mentions

- Already in use
- Top 3 potential

12 February 2015 excl. United States
Model applications with greatest potential (United States)

- Identify prospects more likely to lapse
- Identify prospects more likely to buy
- Improve competitiveness of simplified issue
- Target marketing
- Speed up underwriting process
- Product / underwriting differentiation
- Predict individual mortality
- Identify fraud / misrepresentation
- More competitive preferred premiums
- Predict if applicant has specific diseases

<table>
<thead>
<tr>
<th></th>
<th>2013 survey</th>
<th>2010 survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify prospects more likely to lapse</td>
<td>90%</td>
<td>80%</td>
</tr>
<tr>
<td>Identify prospects more likely to buy</td>
<td>85%</td>
<td>70%</td>
</tr>
<tr>
<td>Improve competitiveness of simplified issue</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>Target marketing</td>
<td>75%</td>
<td>60%</td>
</tr>
<tr>
<td>Speed up underwriting process</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>Product / underwriting differentiation</td>
<td>65%</td>
<td>55%</td>
</tr>
<tr>
<td>Predict individual mortality</td>
<td>60%</td>
<td>50%</td>
</tr>
<tr>
<td>Identify fraud / misrepresentation</td>
<td>55%</td>
<td>45%</td>
</tr>
<tr>
<td>More competitive preferred premiums</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td>Predict if applicant has specific diseases</td>
<td>45%</td>
<td>35%</td>
</tr>
</tbody>
</table>
Model applications with greatest potential (United States)

- Identify prospects more likely to lapse
- Identify prospects more likely to buy
- Improve competitiveness of simplified issue
- Target marketing
- Speed up underwriting process
- Product / underwriting differentiation
- Predict individual mortality
- Identify fraud / misrepresentation
- More competitive preferred premiums
- Predict if applicant has specific diseases

Retention and marketing
Underwriting

12 February 2015
Data
a key ingredient
## Reasons for not using predictive models

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of mentions (excl. U.S.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low business volumes / lack of data</td>
<td>10</td>
</tr>
<tr>
<td>Low priority / cost high vs benefit</td>
<td>6</td>
</tr>
<tr>
<td>Lack of expertise</td>
<td>4</td>
</tr>
<tr>
<td>Lack of resources</td>
<td>3</td>
</tr>
<tr>
<td>Data issues in group business</td>
<td>2</td>
</tr>
<tr>
<td>Data privacy concerns</td>
<td>1</td>
</tr>
<tr>
<td>Not relevant to type of business</td>
<td>1</td>
</tr>
</tbody>
</table>

Among U.S. insurers who had not implemented predictive models, 89% were concerned that there was not enough proof of accuracy.
# Data sources considered

<table>
<thead>
<tr>
<th>Record type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal records from similar insurance activities</td>
<td></td>
</tr>
<tr>
<td>Internal records from elsewhere in the group</td>
<td>banking data</td>
</tr>
<tr>
<td></td>
<td>motor insurance records</td>
</tr>
<tr>
<td>Records provided voluntarily by customers</td>
<td>fitness measurements from wearable technology</td>
</tr>
<tr>
<td>Public records</td>
<td>death registrations</td>
</tr>
<tr>
<td></td>
<td>census data</td>
</tr>
<tr>
<td>Credit records</td>
<td>records purchased from credit scoring agencies</td>
</tr>
<tr>
<td>Other purchased records</td>
<td>consumer classification records</td>
</tr>
</tbody>
</table>
Data sources considered

- Internal records from similar insurance activities
- Public records
- Internal records from elsewhere in the group
- Credit records
- Other purchased records
- Records provided voluntarily by customers

- Currently use
- Likely to use
- Unlikely to use

12 February 2015
Barriers to acquiring data

- No relevant external data has been collected
- Poor historic internal data collection / collation
- Unwillingness of third parties to supply data at all
- Cost of acquiring data from third parties
- Data privacy legislation

Strongly agree | Agree | Disagree | Strongly disagree

12 February 2015
Other barriers flagged

- Consumer attitudes to use of data by insurers
- Previous communications with customers about how their data would be used
- IT constraints including cost, difficulty consolidating data from multiple sources and system discontinuities
- Lack of sufficiently granular data
- Lack of credible data for low frequency events such as death
- Data updated too infrequently
Survey Conclusions
Survey Conclusions

• Implementation rates are currently
  – highest among Australian insurers
  – lowest among Group Business writers (data granularity)

• Most insurers are satisfied with the performance of their implemented models, for the rest it is too soon to tell

• The majority of insurers will be using predictive models within the next couple of years
Survey Conclusions

• Models applications most implemented / with most imminent potential
  – Lapse
  – Propensity to buy
  – Targeted marketing

• Insurers would most like to achieve
  – Underwriting simplification (but not doing away with underwriting entirely)
  – Identification of new risk factors
Obstacles to overcome

• Lack of **available** data
  – Privacy
  – IT constraints
  – Volume, especially for low frequency events
  – Quality

• Lack of expertise and resources

• Cost vs benefit unclear
Examples in action
www.kaggle.com

- “The Home of Data Science”
- Competitions for data problems/predictive modelling
For example…

• Deloitte on Churning

  • “The ability to predict ahead of time when a customer is likely to churn can enable early intervention processes to be put in place, and ultimately a reduction in customer churn. This competition seeks a solution for predicting which current customers of an insurance company will leave in 12 months’ time, and when.”

• 37 Teams

• $70,000 prize
For example…

• Current competition
  – Axa looking at Telematics in cars
  – Data of 50,000 (anonymised!) car trips
  – Driving “signature” – length of journey, acceleration, cornering etc
  – Identify fingerprint of who drove

  – $30,000 prize
The underwriting statistician

Female

Aged 52

Non-smoker, including negative cotinine test

**VERY LARGE SUM ASSURED**

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSYST</td>
<td>100</td>
</tr>
<tr>
<td>BPDIAS</td>
<td>72</td>
</tr>
<tr>
<td>PULSE</td>
<td>72</td>
</tr>
<tr>
<td>ALB</td>
<td>3.6</td>
</tr>
<tr>
<td>ALP</td>
<td>47</td>
</tr>
<tr>
<td>BILI</td>
<td>0.3</td>
</tr>
<tr>
<td>BUN</td>
<td>13</td>
</tr>
<tr>
<td>CHOL</td>
<td>230</td>
</tr>
<tr>
<td>SGLUC</td>
<td>83</td>
</tr>
<tr>
<td>AST</td>
<td>10</td>
</tr>
<tr>
<td>ALT</td>
<td>4</td>
</tr>
<tr>
<td>SPROT</td>
<td>6.1</td>
</tr>
<tr>
<td>TRIG</td>
<td>141</td>
</tr>
<tr>
<td>GLOB</td>
<td>2.5</td>
</tr>
<tr>
<td>HDL</td>
<td>65</td>
</tr>
<tr>
<td>TC/HDL</td>
<td>3.54</td>
</tr>
<tr>
<td>UpH</td>
<td>5.65</td>
</tr>
<tr>
<td>UPROT</td>
<td>13</td>
</tr>
<tr>
<td>UCREAT</td>
<td>237.9</td>
</tr>
<tr>
<td>UGLUC</td>
<td>Neg</td>
</tr>
<tr>
<td>UPROT/CREAT</td>
<td>.055</td>
</tr>
<tr>
<td>ULEUK</td>
<td>Neg</td>
</tr>
<tr>
<td>UHEMO</td>
<td>Neg</td>
</tr>
<tr>
<td>BMI</td>
<td>23.2</td>
</tr>
<tr>
<td>LDL</td>
<td>137</td>
</tr>
<tr>
<td>FRUC</td>
<td>1.4</td>
</tr>
<tr>
<td>LDL/HDL</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Lanzrath, B et al, A Comprehensive Multivariate Approach to the Stratification of Applicant-level All-cause Mortality, ON THE RISK vol.27 n.1 (2011)
Improving risk stratification?

Lanzrath, B et al, A Comprehensive Multivariate Approach to the Stratification of Applicant-level All-cause Mortality, ON THE RISK vol.27 n.1 (2011)

• **Data**
  
  – Insurance application data & laboratory results
    
    6 million applications 2001 - 2008
    
    144 Variables

  – Social Security Death Master File

• **Method**

  1. Link application data to death records to obtain survival estimates

  2. Construct predictive models
    
    Cox Proportional Hazards Multivariate Regression

  3. Rank hazard scores within gender, smoker & age bands
Sample results – no surprises?

Lanzrath, B et al, A Comprehensive Multivariate Approach to the Stratification of Applicant-level All-cause Mortality, ON THE RISK vol.27 n.1 (2011)

![Graph showing ExamOne Applicants and Deaths by Risk Decile: Non-smoking Applicants, 2001-2008]
Sample results – surprise!

Lanzrath, B et al, A Comprehensive Multivariate Approach to the Stratification of Applicant-level All-cause Mortality, ON THE RISK vol.27 n.1 (2011)
Female
Aged 52
Non-smoker, including negative cotinine test

VERY LARGE SUM ASSURED

<table>
<thead>
<tr>
<th>BPSYST</th>
<th>100</th>
<th>SGLUC</th>
<th>83</th>
<th>UpH</th>
<th>5.65</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPDIAS</td>
<td>72</td>
<td>AST</td>
<td>10</td>
<td>UPROT</td>
<td>13</td>
</tr>
<tr>
<td>PULSE</td>
<td>72</td>
<td>ALT</td>
<td>4</td>
<td>UCREAT</td>
<td>237.9</td>
</tr>
<tr>
<td>ALB</td>
<td>3.6</td>
<td>SPROT</td>
<td>6.1</td>
<td>UGLUC</td>
<td>Neg</td>
</tr>
<tr>
<td>ALP</td>
<td>47</td>
<td>TRIG</td>
<td>141</td>
<td>UPROT/CREAT</td>
<td>.055</td>
</tr>
<tr>
<td>BILI</td>
<td>0.3</td>
<td>GLOB</td>
<td>2.5</td>
<td>ULEUK</td>
<td>Neg</td>
</tr>
<tr>
<td>BUN</td>
<td>13</td>
<td>HDL</td>
<td>65</td>
<td>UHEMO</td>
<td>Neg</td>
</tr>
<tr>
<td>CHOL</td>
<td>230</td>
<td>TC/HDL</td>
<td>3.54</td>
<td>BMI</td>
<td>23.2</td>
</tr>
<tr>
<td>SCREAT</td>
<td>0.6</td>
<td>LDL</td>
<td>137</td>
<td>FRUC</td>
<td>1.4</td>
</tr>
<tr>
<td>GGT</td>
<td>55</td>
<td>LDL/HDL</td>
<td>2.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lanzrath, B et al, A Comprehensive Multivariate Approach to the Stratification of Applicant-level All-cause Mortality, ON THE RISK vol.27 n.1 (2011)
What is the Social Security Death Master File?

- Deaths reported to Social Security Administration
  - by hospitals, funeral homes, state offices etc.
- Almost 90 million deaths records added since 1962
  - Name, social security number
- 1980 legal ruling that data must be disclosed
- Widely used in research
  - Cheap subscription rate
  - Weekly & monthly updates
    - More up-to-date than other sources
Issues with the Social Security Death Master File

Bipartisan Budget Act of 2013 Sec. 203

“Restriction on Access to Death Master File”

- Fraud prevention; OR
- Business purpose pursuant to law or fiduciary duty
- Records freely available 3 calendar years after death
One final Predictive Model…

Figure 1: Palm reading example
Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.