Predictive modelling in action:
(Life) lessons from around the world

Adèle Groyer & Alastair Gerrard

12 February 2015

In the fast lane

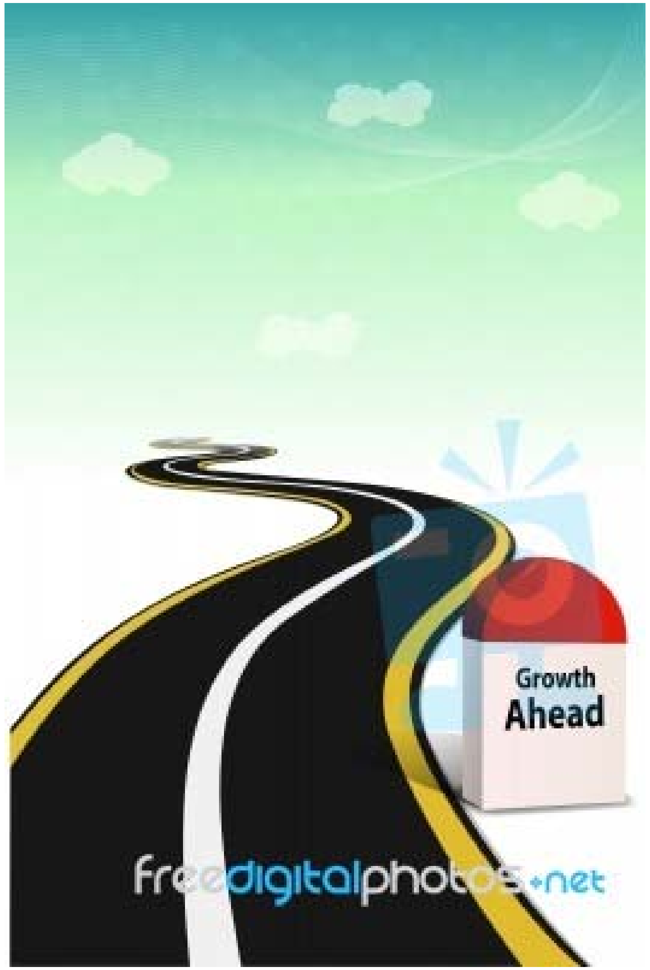
Image courtesy of digidreamgrafix at FreeDigitalPhotos.net

- 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**."

Predictive Modeling for Life Insurance - Ways Life Insurers Can Participate in the Business Analytics Revolution (Deloitte, 2010)

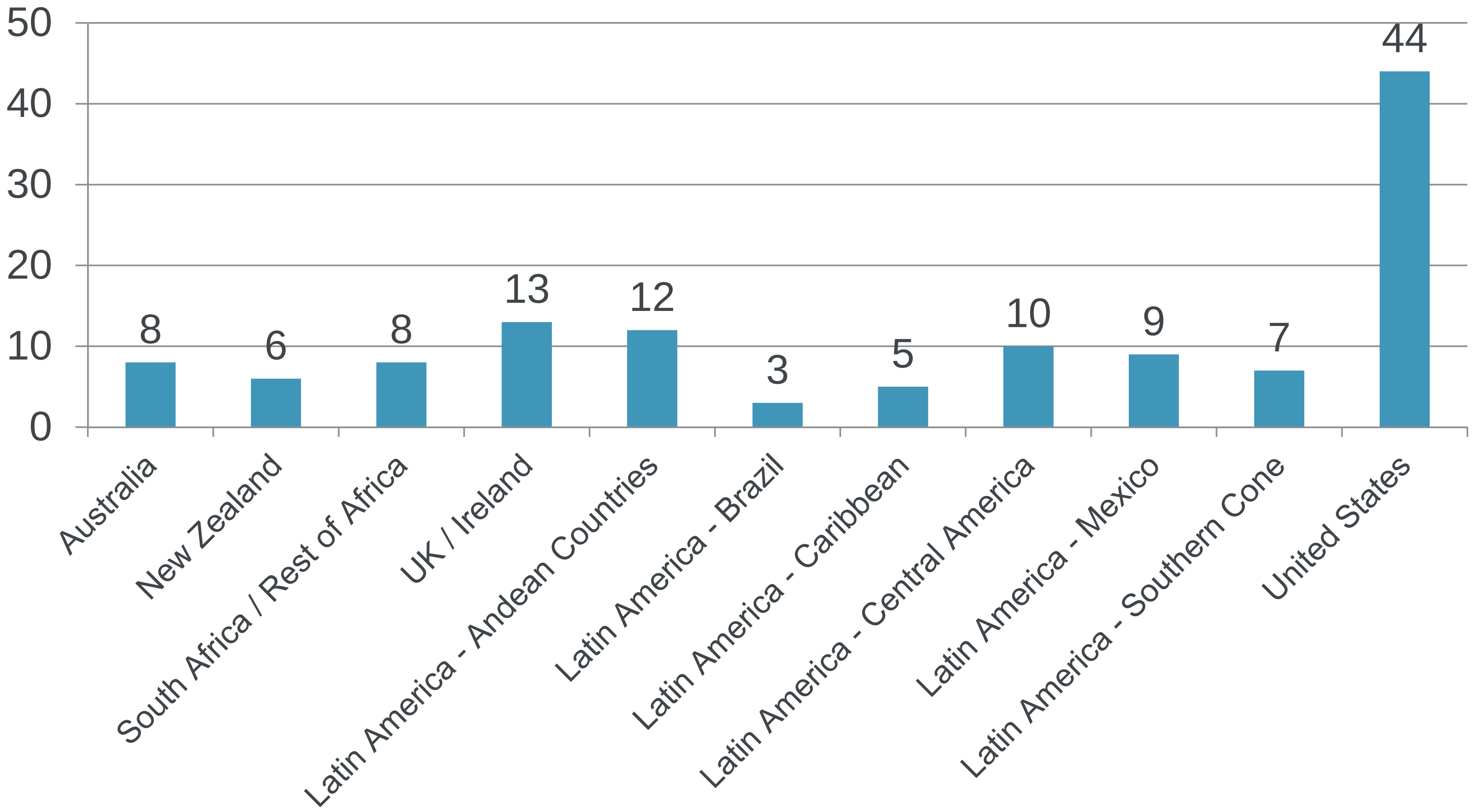
Topics covered by the surveys

- Current and planned use of predictive modelling
- Specific applications in use or with greatest potential use
- Data availability

Institute
and Faculty
of Actuaries

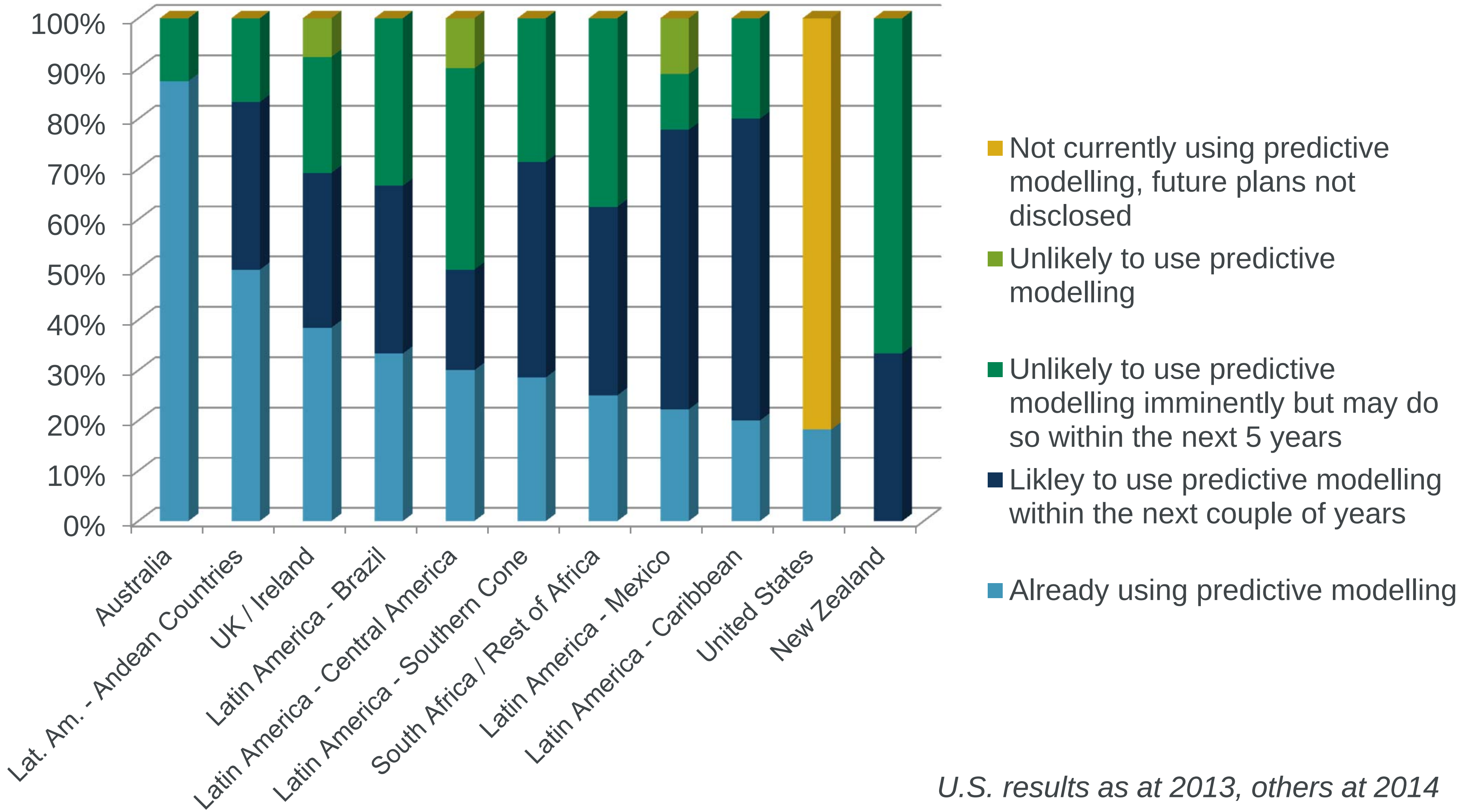
Current and planned use of predictive modelling

Number of responses



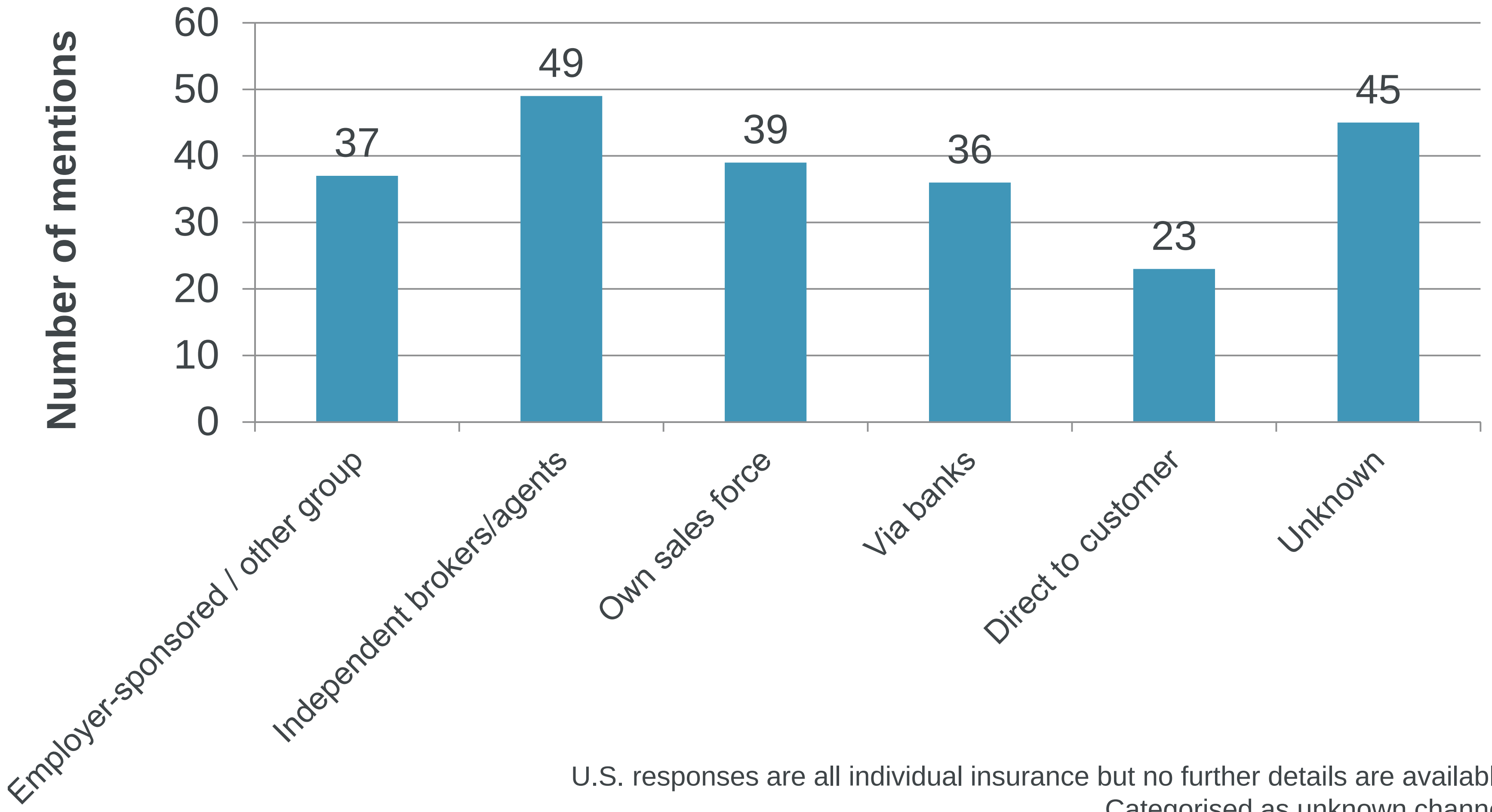
Andean Countries: Colombia, Ecuador, Peru
Caribbean: Puerto Rico, Dominican Republic
Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Panama
Southern Cone: Paraguay, Chile

Current and planned use of models



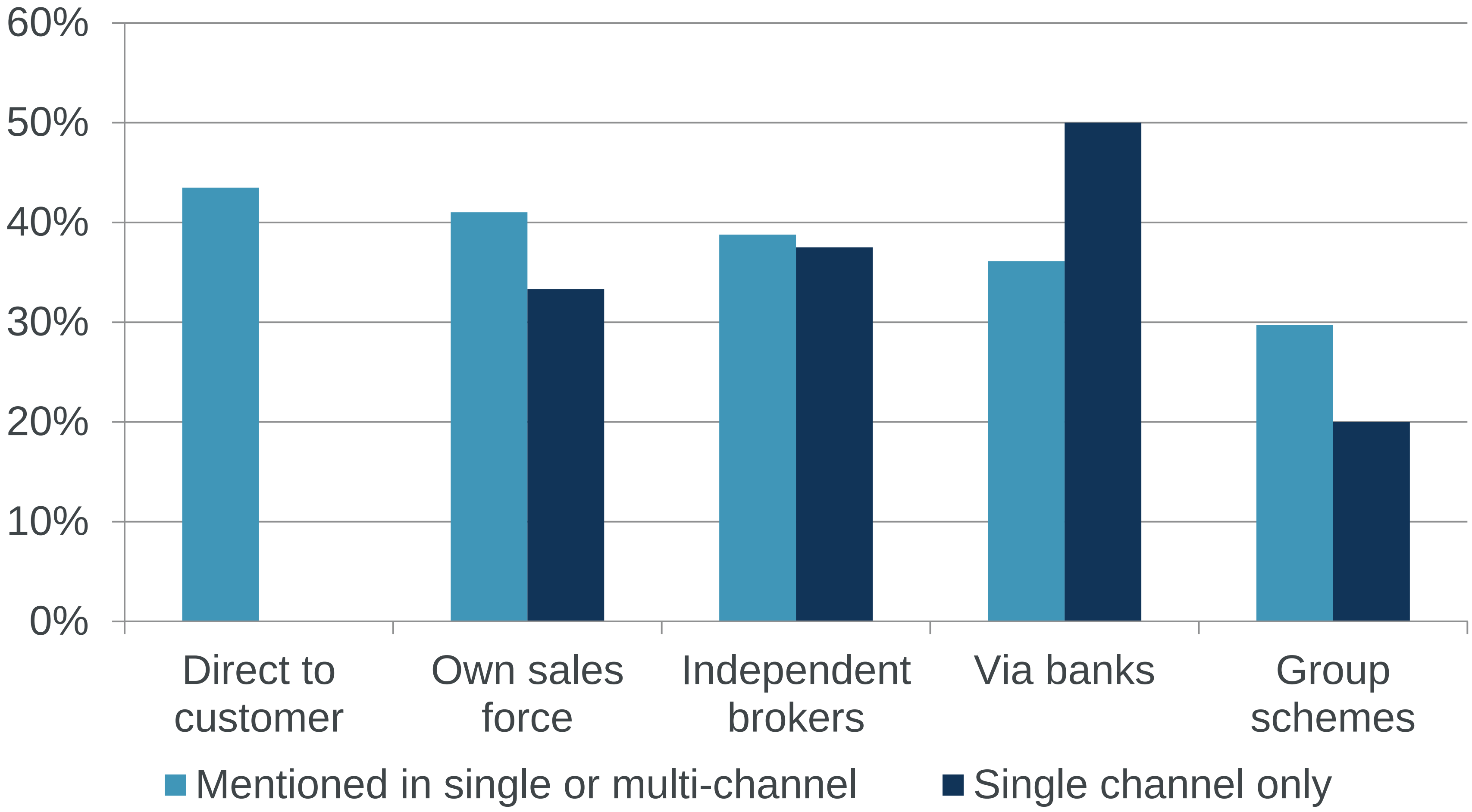
U.S. results as at 2013, others at 2014

Distribution channel



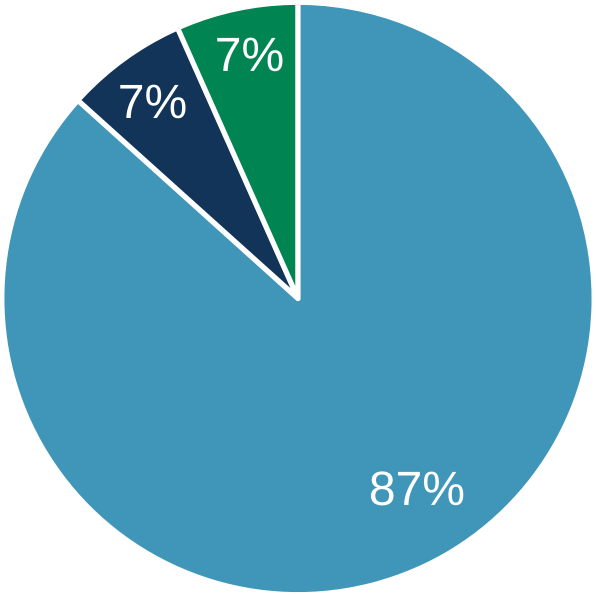
U.S. responses are all individual insurance but no further details are available.
Categorised as unknown channel.

Current model use by channel

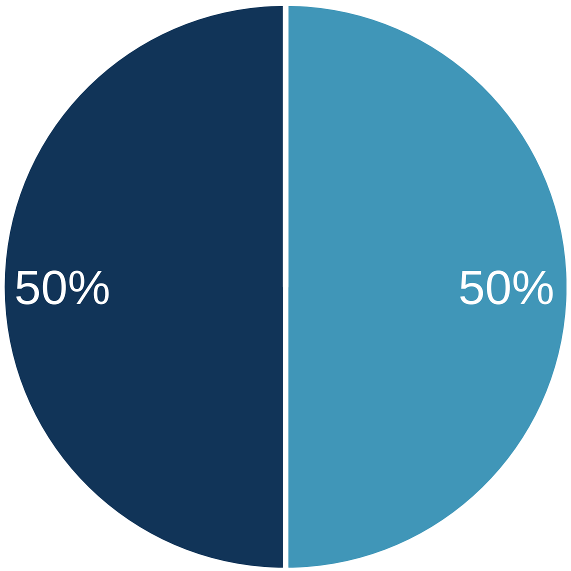


Who developed the model

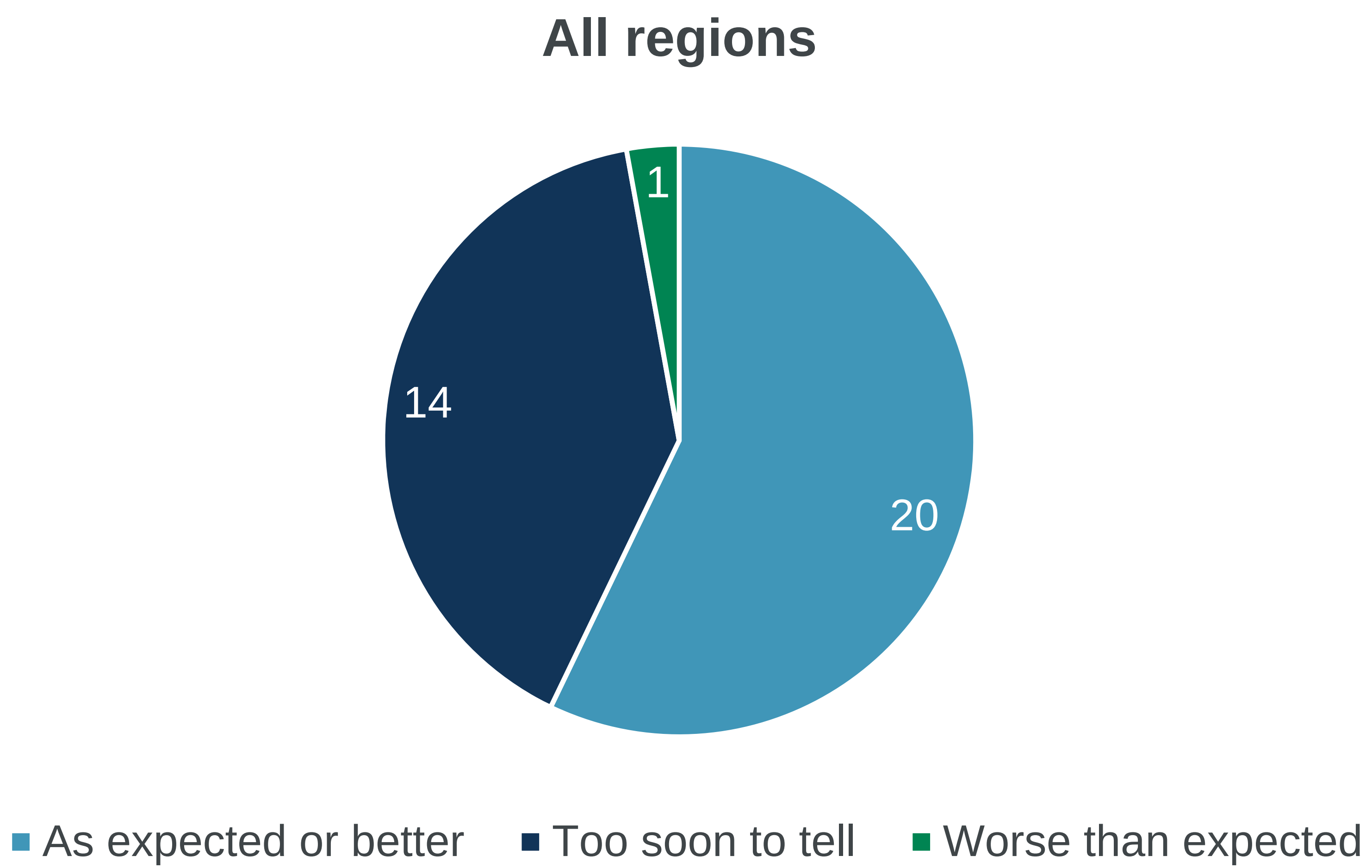
All countries except United States



United States only



Model performance



Reasons for not using predictive models

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

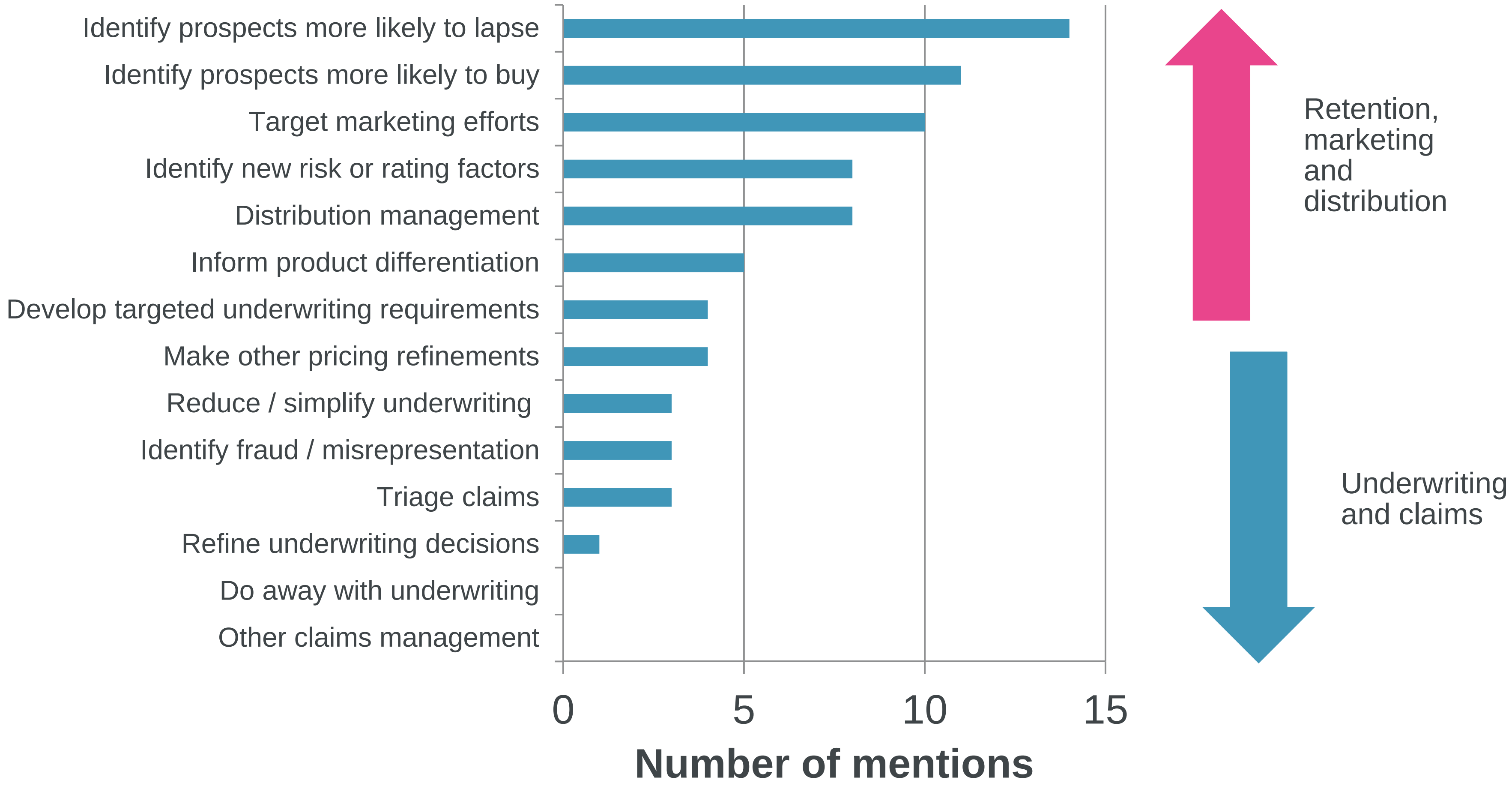
Among U.S. insurers who had not implemented predictive models

- 89% were concerned that there was not enough proof of accuracy
- 33% said it was too expensive

Institute and Faculty of Actuaries

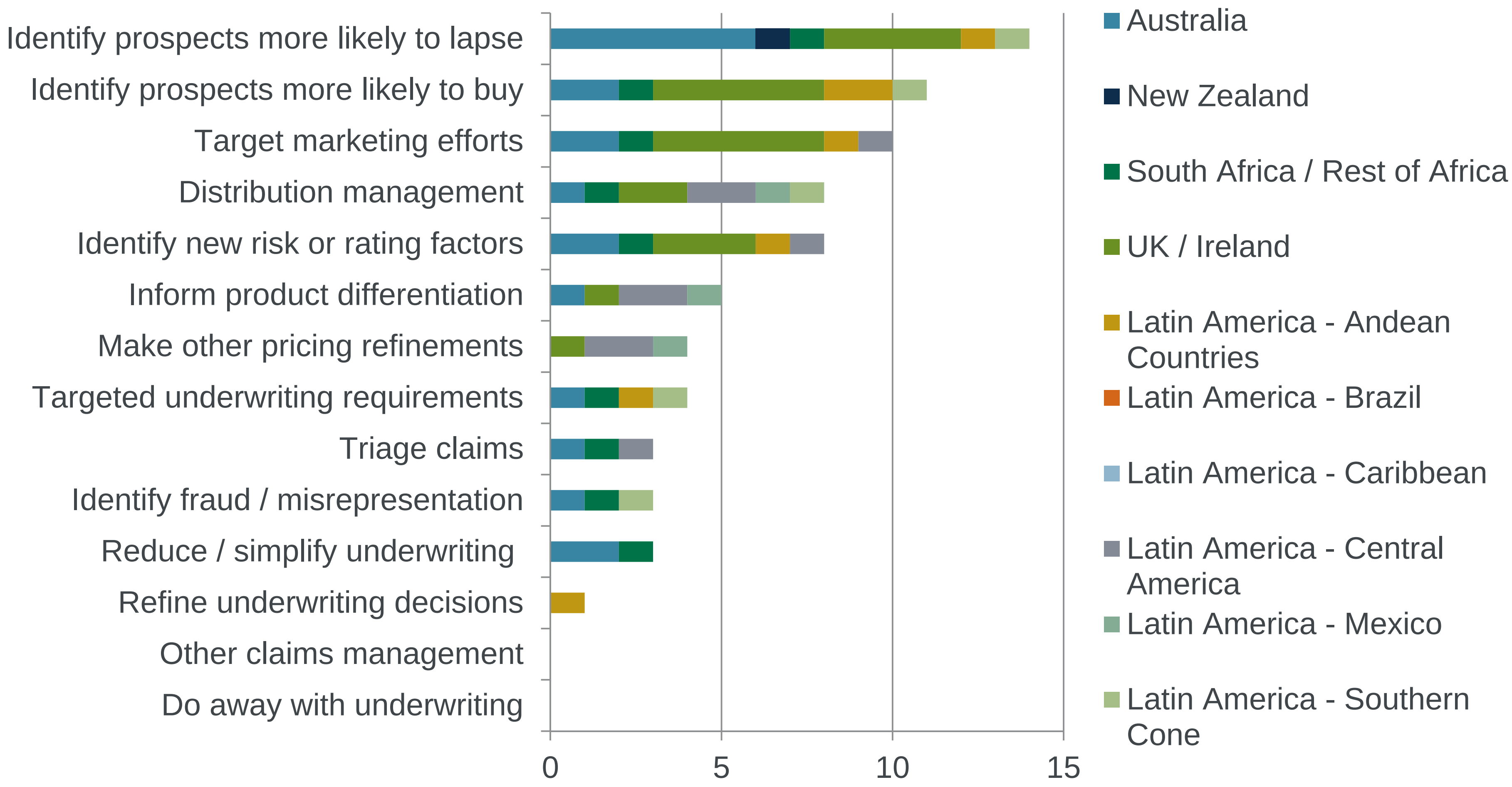
Interest in specific applications

Model applications in use

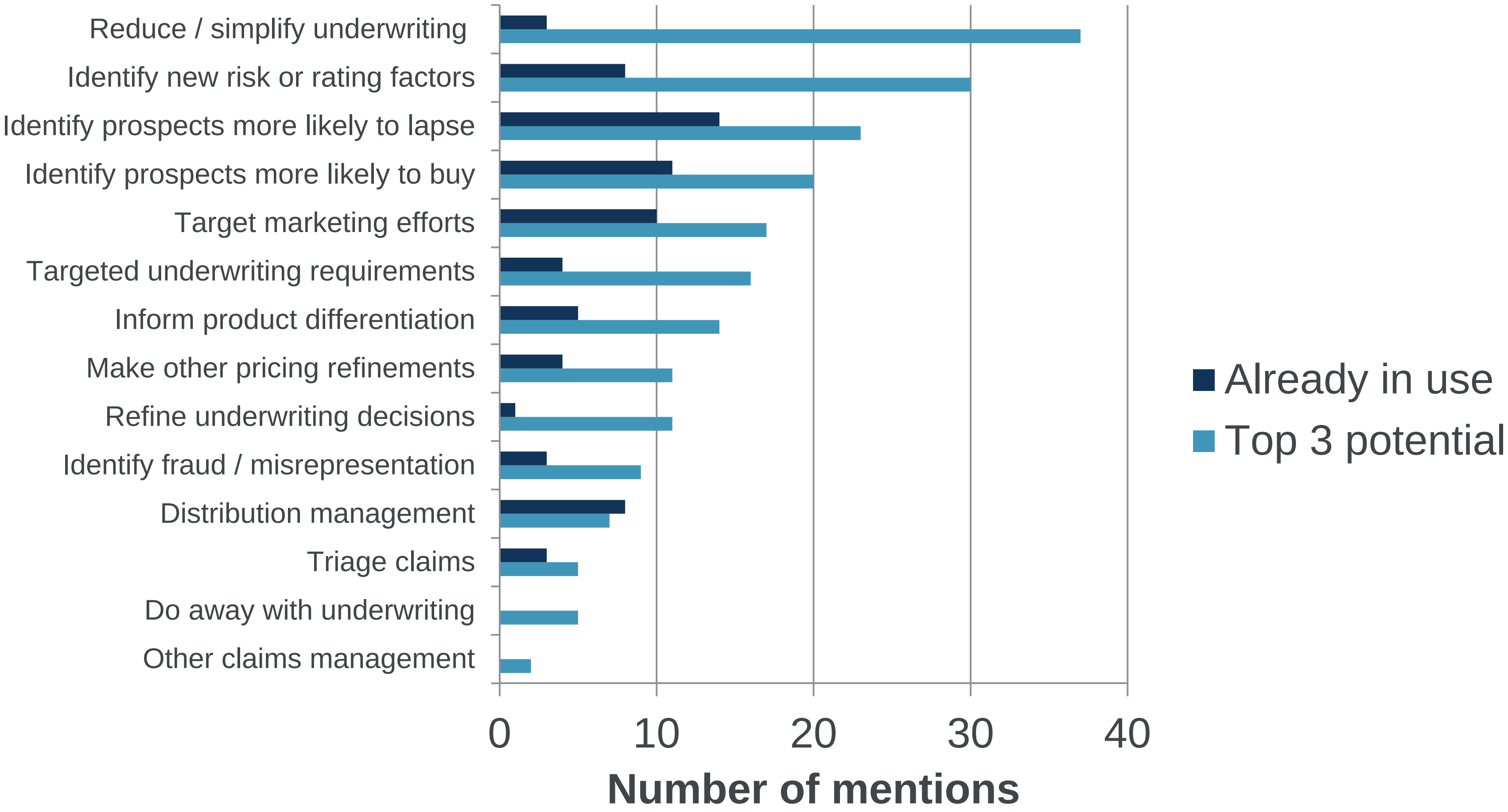


Question not included in U.S. survey

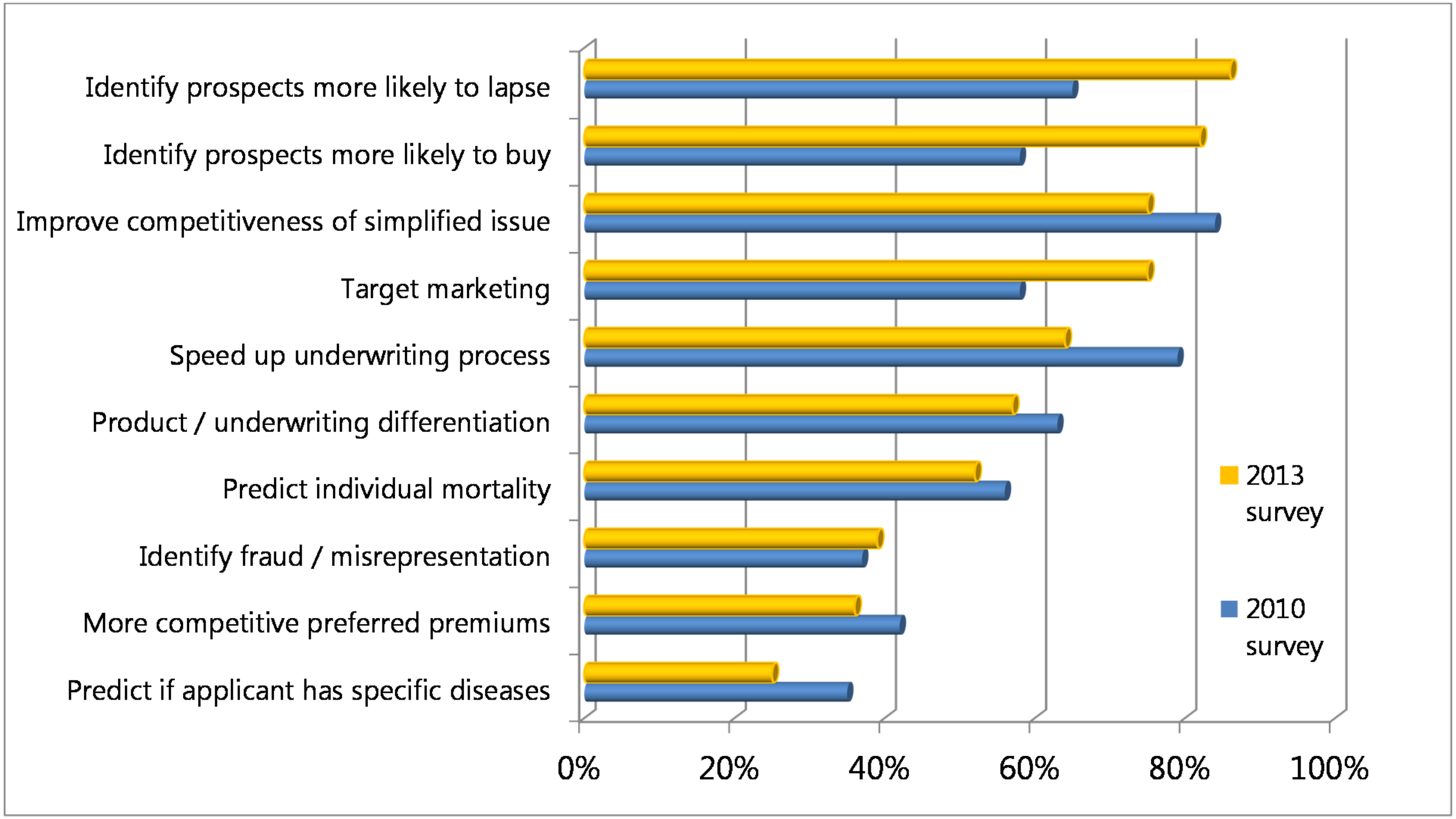
Model application in use by country



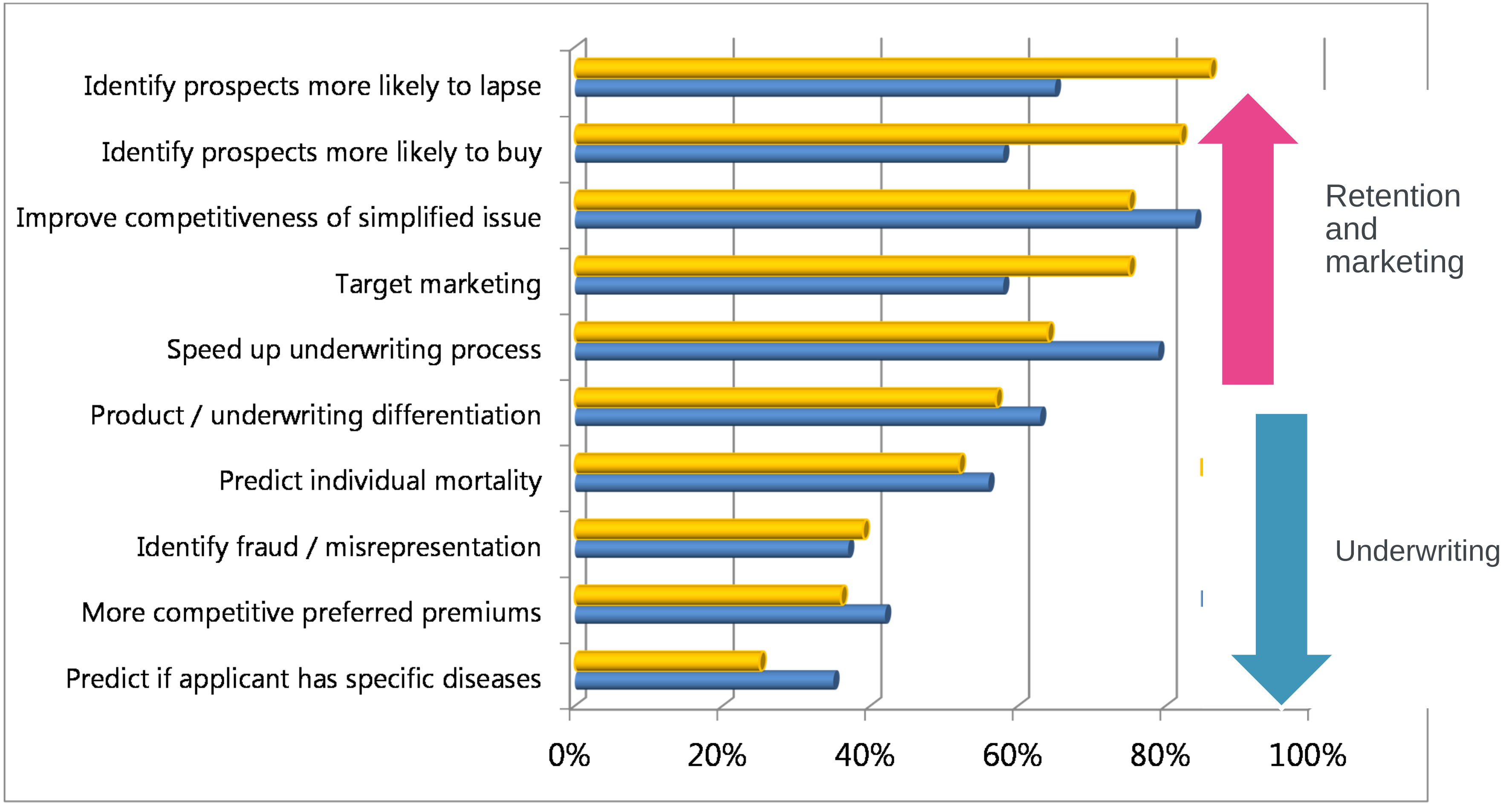
Model applications with greatest potential (excl. United States)



Model applications with greatest potential (United States)



Model applications with greatest potential (United States)



Institute and Faculty of Actuaries

Data

a key ingredient

Reasons for not using predictive models

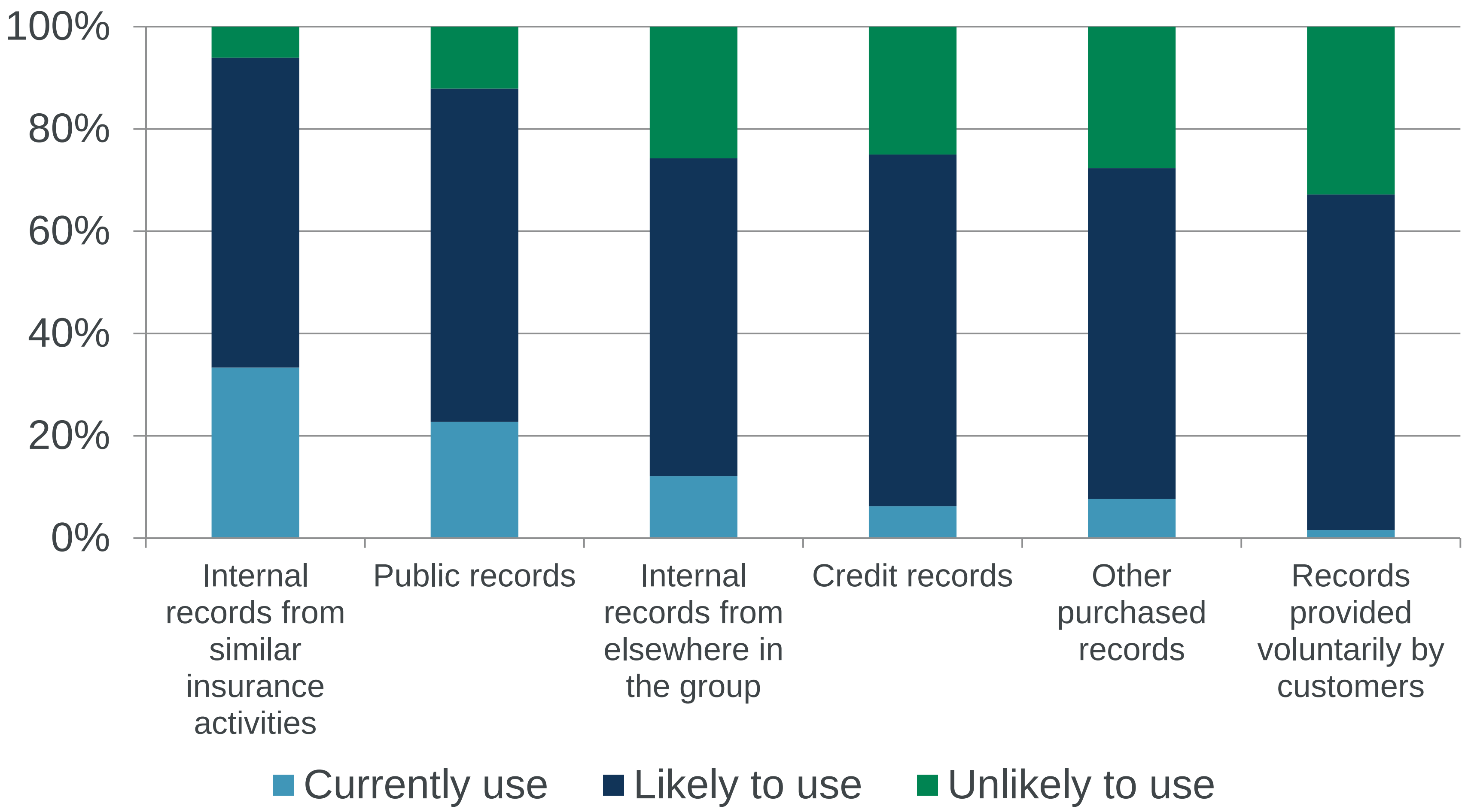
	Number of mentions (excl. U.S.) N = 81
Low business volumes / lack of data	10
Low priority / cost high vs benefit	6
Lack of expertise	4
Lack of resources	3
Data issues in group business	2
Data privacy concerns	1
Not relevant to type of business	1

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

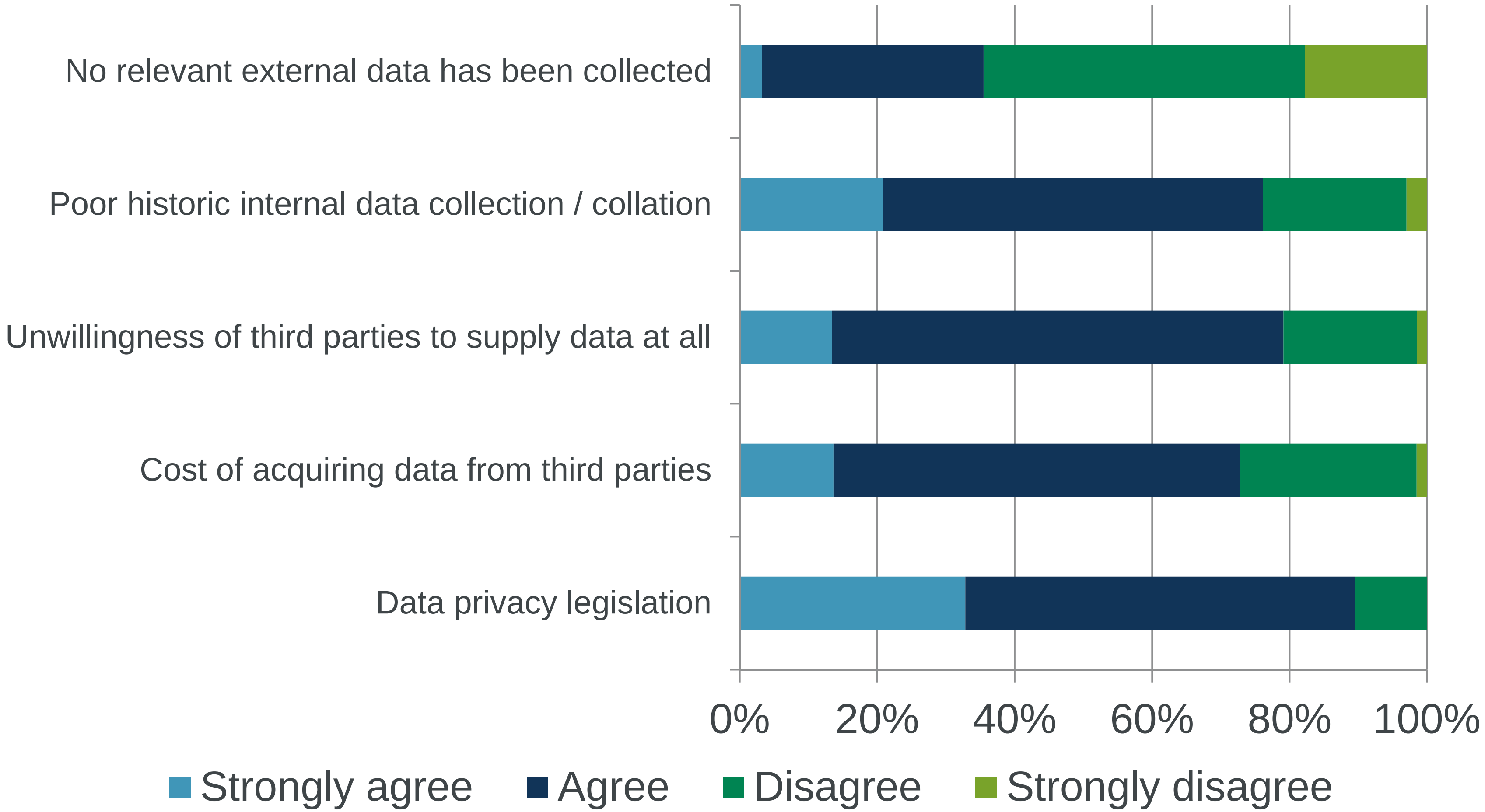
Data sources considered

Record type	Examples
Internal records from similar insurance activities	
Internal records from elsewhere in the group	banking data motor insurance records
Records provided voluntarily by customers	fitness measurements from wearable technology
Public records	death registrations census data
Credit records	records purchased from credit scoring agencies
Other purchased records	consumer classification records

Data sources considered



Barriers to acquiring data



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

Institute
and Faculty
of Actuaries

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

Institute and Faculty of Actuaries

Examples in action

www.kaggle.com

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

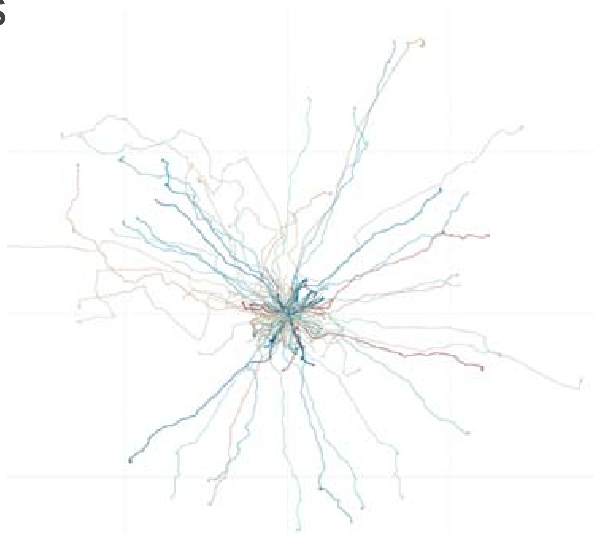
For example…

- 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 (<u>anonymised!</u>) 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

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

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

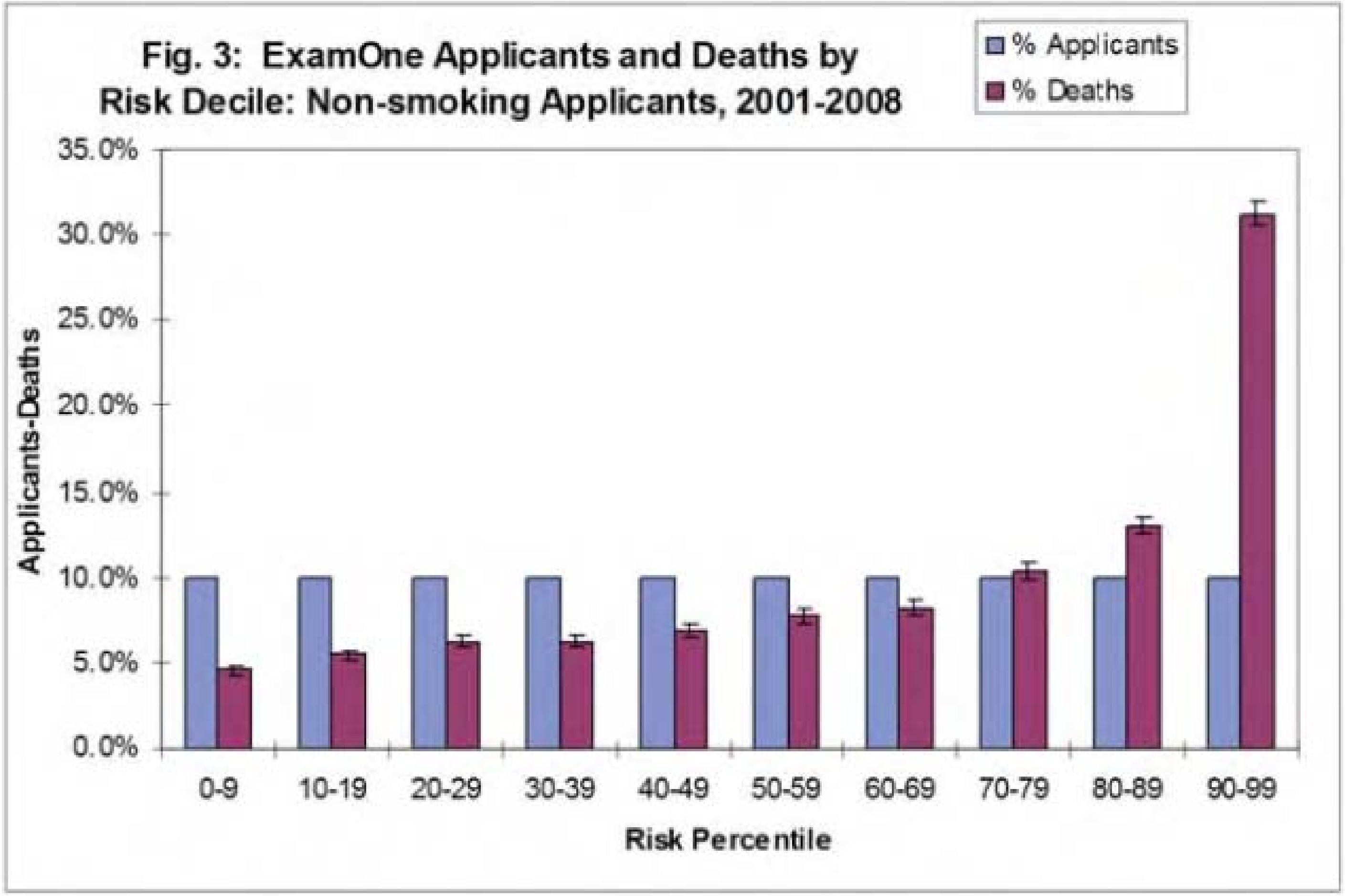
    144 Variables
  - Social Security Death Master File
- **Method**

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

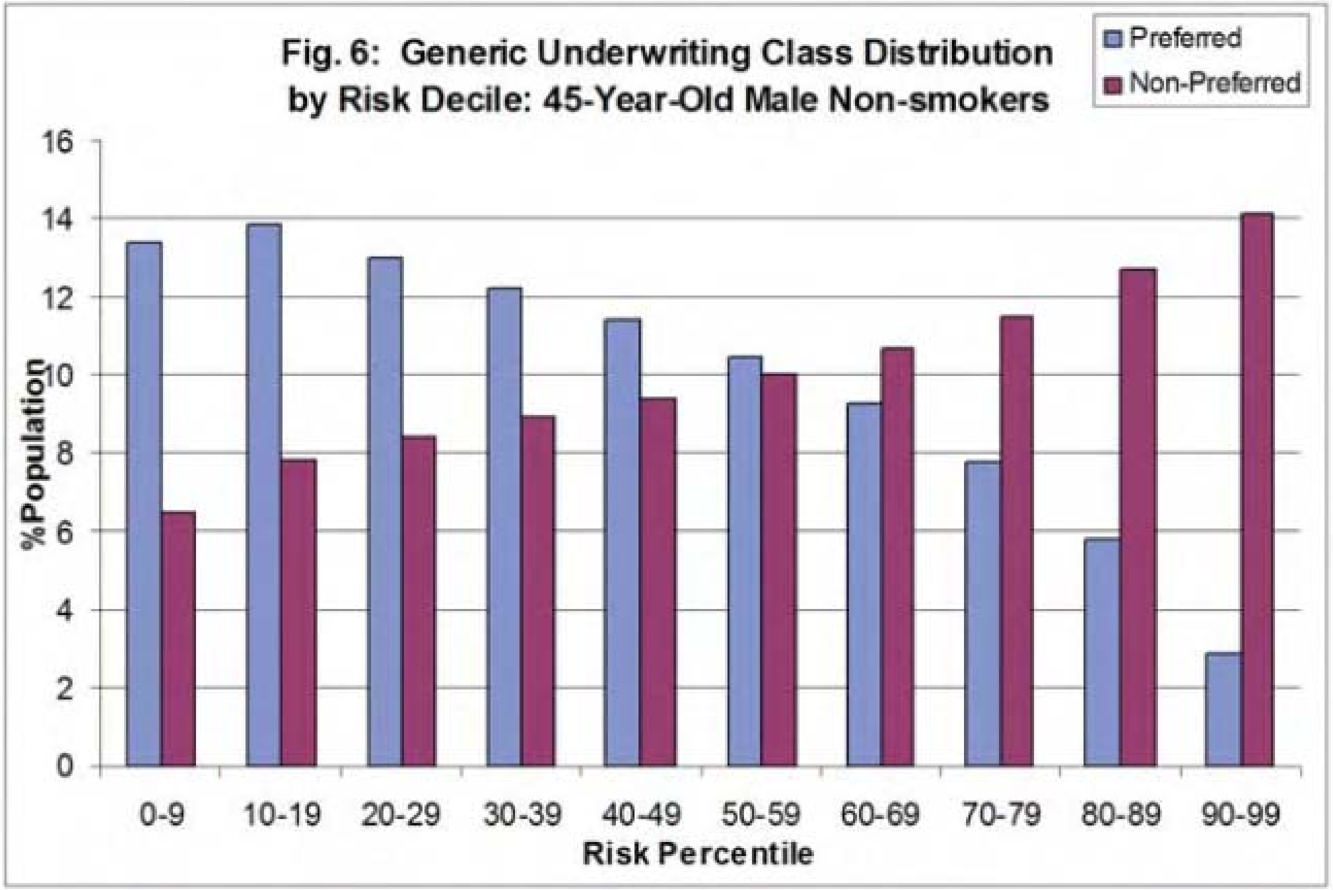
Sample results – no surprises?

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)



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)



Which percentile?

Female

Aged 52

Non-smoker, including negative cotinine test

VERY LARGE SUM ASSURED

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

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…

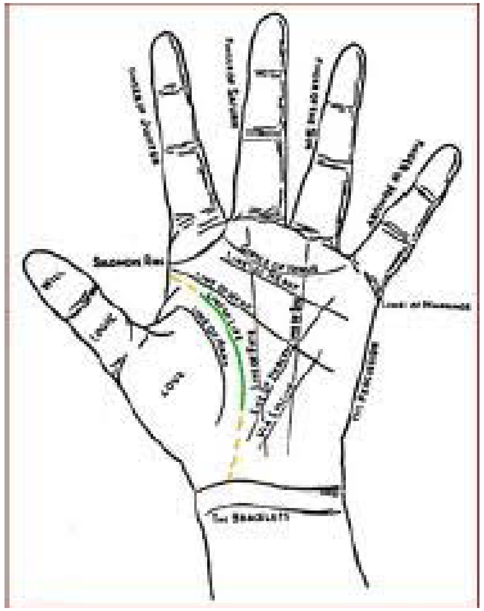
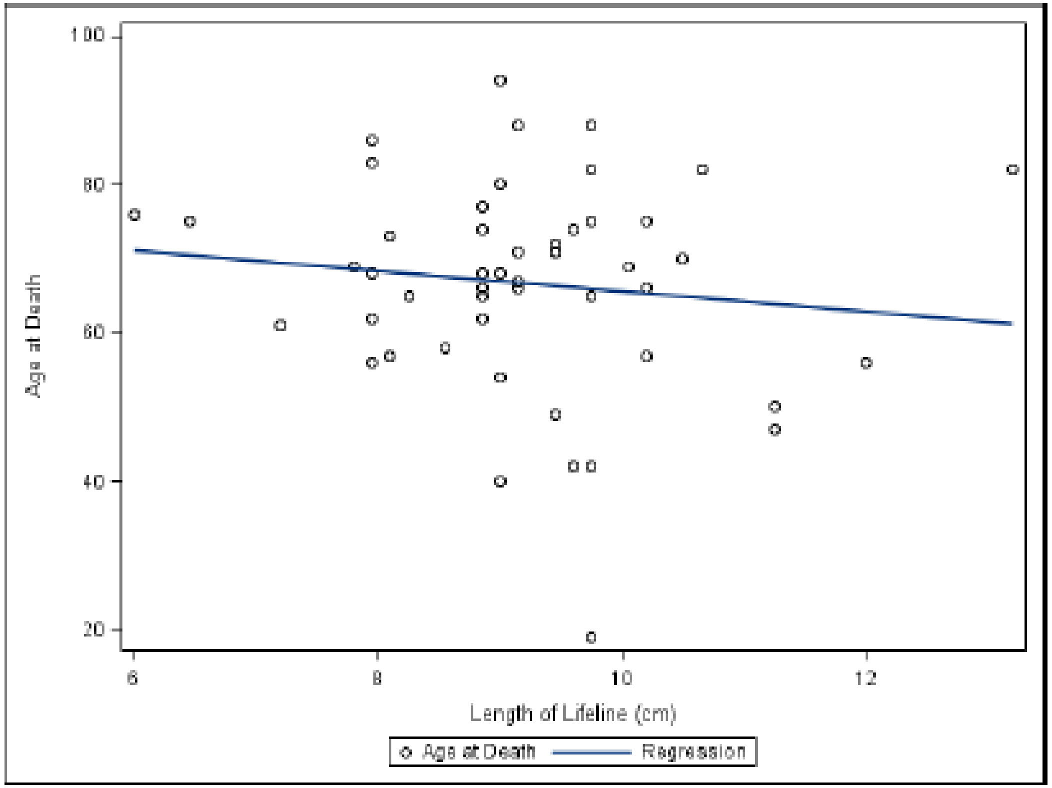


Figure 1: Palm reading example



Questions Comments

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.