



Predictive Modeling, the 21st Century Crystal Ball RGA U.S. Facultative Underwriting

Mark Dion, FALU, FLMI Vice President, Underwriting Rules Development and Education

Wisconsin Association of Health and Life Underwriters (WAHLU) | 4 November 2011



"The Buzz" Wall Street Journal - November 19, 2010

Insurers Test Data Profiles to Identify Risky Clients

"Life insurers are testing an intensely personal new use for the vast dossiers of data being amassed about Americans: predicting people's longevity.

Insurers have long used blood and urine tests to assess people's health—a costly process. Today, however, data-gathering companies have such extensive files on most U.S. consumers—online shopping details, catalog purchases, magazine subscriptions, leisure activities and information from social-networking sites—that some insurers are exploring whether data can reveal nearly as much about a person as a lab analysis of their bodily fluids."



RGA's Position

- RGA does not take a formal position on any particular model or technique
- We respect the power that proper modeling can bring to mortality assessment
- Properly implemented models may have a significant impact on how underwriting will be done in the future – part of a true shift in paradigm



Predictive Modeling

Currently generating a lot of interest. How many models will we need?



Dick Condon - RGA's original predictive "model"



Watson, the jeopardy playing computer is now attempting to diagnose medical conditions



Paul, the football (soccer) predicting octopus, R.I.P.



Not New, but a Relatively "New" Concept for Life Underwriters

- Used in several ways
 - Underwriting
 - Claims
 - Marketing
 - Sales
 - Fraud detection
 - Placement
 - Preferred
 - Super preferred





Predictive Modeling

- A process used in predictive analytics to create a statistical model of future behavior
- A predictive model is a mathematical algorithm that predicts a target variable from a number of factor variables





Predictive Modeling II

- Predictive analytics is the area of data mining concerned with forecasting probabilities and trends
- A predictive model is made up of a number of predictors, variable factors that are likely to influence future behavior of results
- To create a predictive model, data is collected for the relevant predictors, a statistical model is formulated, predictions are made, and the model is validated
- The model may employ a simple linear equation or a complex neural network or genetic algorithm





Predictive Modeling III

- A model is created or chosen to try to best predict the probability of an outcome
- Often the model is chosen on the basis of detection theory
 - Guess the probability of an event given a set amount of input data





Predictive Modeling IV

- Best predictive models depend on rich sets of data from which factor variables can be mined, the model can be built and the model then fitted
- Models range from simple linear regression to advanced techniques, including decision trees, neural networks, generalized linear models and generalized additive models
- The deployment of predictive analytics and models is both art and science.
 Models must be chosen to get the best fit of data and factor variables to produce a strong predictive target outcome
- Bottom line:

Data \rightarrow Model \rightarrow Prediction





Where Did PM Come from?

- Developed by computer scientists, familiar with the methods of if/then/loop analysis
- Massive data-crunching
- Methods of analysis:
 - Look at all the data, using the power of modern computing
 - Instead of using complex mathematical shorthand to reveal trends
 - Primitive, but potentially powerful





Types of Predictive Modeling Tools





Approaches

- Actuarial analyses can be distorted by correlations between unidentified factors
 - This requires a multivariate approach
 - Generalized Linear Modeling is often used
 - Other multivariate approaches can be used, especially if the relationships are non-linear
- Models & Forms
 - Generalized linear models
 - Most frequently encountered for life insurance
 - Neural networks
 - Medicine
 - Genetic algorithms
 - Other Sciences
 - Classification and regression trees (CART or C&RT)
 - Biological classification
 - Census data
 - Others



Modeling methodology

A sample of methods

- Cox Proportional Hazard
- Neural Networks
- Naïve Bayesian Classifiers
- Decision Trees for Classification



Software Tools

- R statistical programming language
 - Perhaps the standard programming language among statisticians for developing statistical software
 - R is widely used for statistical software development and data analysis
- SAS
 - Each SAS program has three major parts:
 - 1. The DATA step
 - 2. Procedure steps (effectively, everything that is not enclosed in a DATA step)
 - 3. A macro language
 - SAS programs allow access to data stored in external data structures and on remote computer platforms
- Data mining
 - Database and spreadsheet tools, from simple to complex



Mathematical Models

A predictive model is a process to derive the value of Y, where:

 $Y = (y_1, y_2, ..., y_N) \text{ from } \{ x_{i1}, x_{i2}, ..., x_{ip} \}$

Based on

 $y_i = f\{ x_{i1}, x_{i2}, ..., x_{ip} \}$

A traditional model form is:

 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{1p} x_{ip}$



Generalized Linear Models (GLM)

- Linear regression analysis
- In statistics, the generalized linear model (GLM) is a flexible generalization of ordinary least squares regression







Cox Proportional Hazard Model

$$\begin{split} \log h_i(t) &= \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{ik} + + \beta_k x_{ik} \\ & \text{or, equivalently,} \\ h_i(t) &= h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{ik} + + \beta_k x_{ik}) \end{split}$$





Predictive Models the Miracle



"I think you should be more explicit here in step two."



Building a Model

Predictive Model Build Process





Model Validation

- Data subset is held back (Holdout Testing)
- Run the model against this reserved data set to see if the model holds true
- Provides an unbiased measure of performance, assuming the test data is used only once to test the model
 - If used more than once the measure cannot truly be unbiased
- The Tuning Set Additional testing another set of data is needed to make adjustments to the model, to experiment with different numbers of predictors, etc. This set must also be randomly assigned.



Predictive Modeling Outside Life Insurance

Outside the Insurance Industry

- Marketing
- Dating services
- Auto insurance rates
- Health care utilization
- Medical treatment
- Modern Examples
 - Amazon
 - Netflix
 - Match.com
 - Anvita Health
 - TSA
 - Watson the Jeopardy playing computer



Examples of Predictive Modeling - Amazon.com

- We determine your interests by examining the
 - Items you've purchased
 - Items you've told us you own
 - Items you've rated
- We then compare your activity on our site with that of other customers, and using this comparison, are able to recommend other items that may interest you



Examples of Predictive Modeling - Netflix

- CineMatch is a database that uses information from three sources to determine which movies customers are likely to enjoy:
 - The films themselves, which are arranged as groups of common movies
 - The customers' ratings, rented movies and current queue
 - The combined ratings of all Netflix users
- According to Netflix, 60 percent of subscribers add these suggested movies to their queues
- Algorithm improved through the Netflix prize competition



Examples of Predictive Modeling - Match.com

- Codenamed "Synapse", the Match algorithm uses a variety of factors to suggest possible mates
- User's stated preferences
 - Desired age range
 - Hair color
 - Body type
- It also learns from their actions on the site
 - "if a woman says she doesn't want to date anyone older than 26, but often looks at profiles of thirty-somethings, Match will know she is in fact open to meeting older men."
 - "the politics one is quite interesting. Conservatives are far more open to reaching out to someone with a different point of view than a liberal is."
- Synapse also uses "triangulation". That is, the algorithm looks at the behavior of similar users and factors in that information, too



Examples of Predictive Modeling - Anvita Health

- Health care analytics company serving doctors, health plans, pharmacy benefit managers, disease management companies, point-of-care IT systems, personal health record (PHR) providers, and other clinical providers
- Provides the analytics engine for Google Health.
- Founded by physicians in 2000 as SafeMed to help prevent avoidable medical errors
- Its name change in 2009 to Anvita Health signaled an expansion of the company's focus beyond drug safety to also include point-of-care clinical decision support, gaps in care analysis, prioritization of patient interventions, predictive modeling, etc.



Examples of Predictive Modeling - TSA





- "A behind-the-scenes watch list matching process that vets passengers against government watch lists."
- "Secure Flight makes travel safer by more effectively identifying individuals that may pose a known or suspected threat to aviation."
- Details are obviously difficult to come by, but risk factors may include:
 - Name, compared to No Fly List
 - Amount of baggage checked relative to length of flight
 - Seat chosen on flight
 - Amount of time in advance flight was booked.
 - Personal profile information (race? religion?)



Examples of Predictive Modeling - Watson, IBM's Jeopardy Playing Computer

- DeepQA Technology
- A way to look at unstructured data, natural language
- Read countless documents and remember them
- Watson can see interdependencies
- "Confidence"





Predictive Modeling in Insurance

- Many potential applications
 - Target Marketing
 - Likelihood to buy
 - Likelihood to persist
 - Likelihood to claim
 - Risk Selection / Underwriting
 - Retention management
 - Fraud Detection





Using Predictive Models in Life Insurance

- Marketing
- Pricing
 - Lapse assumptions
 - Reinsurance pricing
- Claims
 - IBNR (Incurred but not reported)
 - Fraud detection (?!)
- Underwriting
 - Preferred
 - APS or other requirement ordering
 - Fraud detection (?!)
- Focus of most attention is on the impact for life underwriting
 - Lab scoring
 - Models using lifestyle-based factors



Vintage Life Underwriter?



Specifically PM May Help in the US Middle Market First

For the middle market underwriting, examples

- 1. Surrogate for fluid testing
 - or as an adjunct
- 2. Identify preferred issue client profiles
- 3. Labs, BMI, Gender, Smoking status to provide a relative score
- 4. Triage there is promise...however...
- 5. Target marketing
- 6. "Smarter" APS ordering

Haven't we heard this before?



Current PM Activity in Life/Health Underwriting

- The information on the following slides comes from publically available sources including the vendor's websites, publications, and public presentations
- Comments in quotation marks are directly from the vendor's sources
- Neither openly endorsing nor disparaging any of these vendor solution
- None of these companies are affiliated with RGA, although we have had working relationships with all companies on different projects
- Companies should come to their own conclusions about the effectiveness and appropriateness of any predictive modeling applications



Current PM Activity in Life/Health Underwriting BioSignia

Vendor/Product

 BioSignia is a science & technology company with an underwriting product called Mortality Assessment Technology (MAT)

Explanatory Variables

- "Data normally found on the insurance application" captured through data entry or electronically
- Calibration
 - "Incorporates and continuously updates all relevant medical studies"
 - Model has been validated in an independent actuarial mortality study (performed by RGA)

Implementation

- "Enables life insurance professionals to better classify an individual's mortality into standard, preferred, and super-preferred categories more confidently."
- "Accepts and transmits data, easily and electronically, via the internet"
- "Requires only minimal installation and training (SaaS technology)"

RGA®

Current PM Activity in Life/Health Underwriting ExamOne - RiskIQ

Vendor/Product

 ExamOne is a major provider of lab analysis of fluid profiles in life insurance applications with a PM product called RiskIQ

Explanatory Variables

 "Risk IQ draws from more than 140 different variables – ranging from laboratory results to age – to provide a score that ranks a life insurance applicant's risk for mortality."

Calibration

 The model is calibrated to data from millions of life applicants, their fluid profile and the Social Security Death Master File (SSDMF)

Implementation

 RiskIQ is a unique score between 0 and 99 representing the prediction of the proportion of applicants that have a better relative mortality to the applicant within the total cohort. A score of 37 means 37% of applicants in a cohort (age/sex) have better relative mortality than the applicant with a score of 37

RGA®

Current PM Activity in Life/Health Underwriting CRL - SmartScore

Vendor/Product

 CRL is a major vendor of lab analysis and have developed a predictive scoring model called SmartScore

Explanatory Variables

Lab tests and paramedical information

Calibration

 The model is calibrated to data from millions of life applicants, their fluid profile and the SSDMF

Implementation

- "Provides both a composite score and a component score for each test or measurement so the cause(s) of an elevated total score can be quickly and precisely identified"
- "Discriminates risk from the best 10% to worst 10% of applicants in near linear manner so effective not just for standard vs. substandard but between preferred classes"



Current PM Activity in Life/Health Underwriting Heritage Labs

Vendor/Product

 Heritage Labs is a major vendor of lab analysis and is developing a "Risk Score" system

Explanatory Variables

Lab tests and paramedical information

Calibration

 The model is calibrated to data from millions of life applicants, their fluid profile and the SSDMF

Implementation

- Independent scores for liver function, renal function, etc. are provided and rolled into a overall score
- The scores will reflect an all-cause mortality risk as a percentage of a standard accepted mortality table



Current PM Activity in Life/Health Underwriting Deloitte

Vendor/Product

 Deloitte is a global consulting firm and major provider of predictive models for Property & Casualty insurance

Explanatory Variables

- Application, MIB, MVR
- Third party marketing data (e.g. Equifax)
- Does not include lab values

Calibration

 Calibrated to underwriting decisions for cases that went through full medical underwriting

Implementation

 Process facilitates a triage approach whereby best risks from the model may be able to avoid fluid testing



Predictive Modeling Touches What We Do

- Pricing
 - Lapse assumptions
 - Reinsurance pricing
- Marketing
- Claims
 - Incurred But Not Reported (IBNR)
- Underwriting
 - Preferred
 - APS or other requirement ordering
- Placement (RGA for example)
- Fraud detection
 - Underwriting, Claims, Other Financial Fraud
 - Why isn't this a priority?



Risks in Predictive Modeling

- PM is usually retrospective, looking backward
- Behavior is subject to change
- Knowledge of the domain in question is critical
- Models are software based and subject to bugs, and miscalculations
- Data can be biased, incomplete, or simply wrong



Predictive Modeling Controversies



"There is a dark side to electronic records"

- Hank George, Best's Review, August 2010

"Would you buy a life insurance policy from this machine"

- Wall Street Journal, March 12, 2011

 "Credit data is behavioral in nature and...bears no direct causal relationship to insurance loses...It is a natural leap to consider other sources...such as lifestyle, purchasing, household, social network and environmental data."

- Predictive Modeling for Life Insurance. Deloitte, April 2010

 "A perceived 'black box' nature, makes it difficult to describe and explain results; the proprietary nature and structure of each model reinforces the perception."

> - *Predictive Modeling, A Life Underwriter's Primer*, Mark Dion, On The Risk, volume 27, n.2. 2011



Vendor PM Questionnaire

- General Questions
 - Who developed it?
 - What does it model, etc.
- Implementation
- Data & Variables
- Modeling Approach and Validation
- Maintenance
- Liability
- Future Plans



Email me for a full list: mdion@rgare.com



Consider the Unresolved Questions

- Mortality implications
- Regulatory concerns
- Over-fitting
- What about using multiple models?
 - Overlapping data
 - Conflicting data
 - Reinforcement ? Or contradiction?
- Ultimately, do the models fit with our current processes?
- Or does their presence in the marketplace finally move us to a new paradigm of risk selection?



Our Role Before Implementation

Questions Underwriters Should Consider

- Methodology for validation for your business
- What is the model trying to predict?
 - Behavior?
 - Outcome?
 - Effectiveness?
- Who should assist in the assessment?
 - Underwriters & medical directors
 - Actuarial, pricing, valuation
 - Compliance
 - Sales, marketing, business development
 - Who else?
- Protective Value
- Cost and Benefit
- Exclusivity





Report of the Society of Actuaries Predictive Modeling Survey Subcommittee

Survey Sections

- Executive Summary
- Demographics
- Marketing
- Underwriting
- Reinsurance
- Claims
- Risk Mitigation
- Regulatory Issues
- Market Conduct



Society of Actuaries 475 N. Martingale Rd., Ste. 600 Schaumburg, IL 60173 Phone: 847-706-3500 Fax: 847-706-3599 Web site: http://www.soa.org

Completed and pending peer review prior to publication



Effects of Implementation and Deployment

- "Better" risk assessment does not, in itself, lead to lower aggregate mortality
 - Often we simply move individuals between classes without actually improving mortality results of the product
- New underwriting tools may lead to
 - Different placement ratios
 - Different risk pool of applicants
 - Field force selection effects
 - Early adopter selection effects
 - Change in profitability models
 - Variations in understanding in how the tool should work



Future State of Underwriting

- Predictive models will become more common in life insurance
- Ten years from now they will be simply another tool in the underwriters tool box
- Prepare yourselves and mentor newer underwriters to understand these tools as we would any other...





Thoughts on the Future

Remember: Underwriting exists to solve an *information asymmetry* problem between buyers and sellers of insurance

In the future:

More information available to applicants

- Home tests/kiosks for specific diseases (HIV, diabetes, etc.)
- Genetic testing
- Internet "self-diagnosis"

Less information available to insurers

- Increased privacy/consumer regulation
- Move toward simplified issue underwriting





More Thoughts on the Future

Data, data, and more data



- The amount of data will continue to increase at an exponential pace
- Strong temptation to utilize every available piece of data to solve or bridge the information asymmetry gap
- We must be smarter about how we filter through the data to separate out the signal from the noise
- Hank George: playing "Russian roulette with our right to underwrite"
- We must be very careful to uphold the public trust by using data appropriately



Questions





Further Reading

- Bellazzi, Riccardo, Zupan, Blaz; Predictive data mining in clinical medicine: Current issues and guidelines; Int. J. Med. Inform. 2006 doi:10.1016/j.ijmedinf.2006.11.006
- Ciardello, Gary, McLeroy, David; Predictive Modeling's New Lease on Life; Actuarial Software Now, American Academy of Actuaries, Winter 2011, pp 24-28
- Draagghtel, Kesnia; Predictive Modeling with Consumer Data; The Actuary, Oct/Nov 2011 pp 28-37
- Vickers PhD, Andrew J., Prediction Models in Cancer Care; CA Cancer J Clin 2011; 61:315-326
- Wesley MD, David, Cox, Hugh, Modeling Total Cholesterol as a Predictor of Mortality, The Low-Cholesterol Paradox; Journal of Insurance Medicine. J Insur Med 2011; 42:62-75

Sources and Acknowledgements

- Berry, Michael J. A. and Linoff, Gordon; Data Mining Techniques for Marketing, Sales and Customer Management; John Wiley and Sons, Inc; 2004
- Cox D.R., *Regression Models and Life Tables*, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 34, No. 2, 1972 pp. 187-220
- Galen, Robert S., Gambino, S. Raymond; *Beyond Normality: the Predictive Value and Efficacy of Medical Diagnoses*; John Wiley and Sons, Inc; 2001
- Geisser, Seymour; Predictive Inference: An Introduction. Monographs on Statistics and Applied Probability 55; New York: Chapman & Hall, 1993
- Report of the Society of Actuaries Predictive Modeling Survey Subcommittee, prepublication
- Rozar, Tim; Shining a Light in the Black Box, 2011 Midwest Underwriting Conference presentation
- Weiss, Sholom M. and Kulikowski, Casimir A.; Computer Systems That Learn, Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems; Morgan Kaufman Publishers, Inc., 1991
- Special thanks to Rodney Brown, Abe Gootzeit, Dave Wheeler





Predictive Modeling, the 21st Century Crystal Ball RGA U.S. Facultative Underwriting

Mark Dion, FALU, FLMI Vice President, Underwriting Rules Development and Education

Wisconsin Association of Health and Life Underwriters (WAHLU) | 4 November 2011