Article from
Product Matters
February 2018
Issue 109

Predictive Modeling for Life Insurance

Ways Life Insurers Can Participate in the Business Analytics Revolution

By Chris Stehno and Jim Guszcza
Contributors: Mike Batty, Arun Tripathi, Alice Kroll, Cheng-sheng Peter Wu, David Moore, and Mitch Katcher

Editor's Note: This article is part one of two. It has been formatted for the newsletter and is reprinted with permission from Deloitte Development LLC. Stay tuned for Deloitte's next release on predictive modeling, which will include some newer concepts and solutions such as Electronic Health Records and Behavioral Health.

THE RISE OF "ANALYTIC" DECISION MAKING

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 [1,2]. 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. In a sense, this is a discipline that actuaries have practiced for quite a long time. Indeed, one of the oldest examples of statistical analysis guiding business decisions is the use of mortality tables to price annuities and life insurance policies (which originated in the work of John Graunt and Edmund Halley in the 17th century). Likewise, throughout much of the 20th century, general insurance actuaries have either implicitly or explicitly used Generalized Linear Models [3,4,5] and Empirical Bayes (a.k.a. credibility) techniques [6,7] for the pricing of short-term insurance policies. Therefore, predictive models are in a sense, "old news." Yet in recent years, the power of statistical analysis for solving business problems and improving business processes has entered popular consciousness and become a fixture in the business press. "Analytics," as the field has come to be known, now takes on a striking variety of forms in an impressive array of business and other domains.

Credit scoring is the classic example of predictive modeling in the modern sense of "business analytics." Credit scores were initially developed to more accurately and economically underwrite and determine interest rates for home loans. Personal auto and home insurers subsequently began using credit scores to improve their selection and pricing of personal auto and home risks [8,9]. It is worth noting that one of the more significant analytical innovations in personal property-casualty insurance in recent decades originated outside the actuarial disciplines. Still more recently, U.S. insurers have widely adopted scoring models—often containing commercial credit information—for pricing and underwriting complex and heterogeneous commercial insurance risks [10].

The use of credit and other scoring models represents a subtle shift in actuarial practice. This shift has two related aspects. First, credit data is behavioral in nature and, unlike most traditional rating variables, bears no direct causal relationship to insurance losses. Rather, it most likely serves as a proxy measure for non-observable, latent variables such as "risk-seeking temperament" or "careful personality" that are not captured by more traditional insurance rating dimensions. From here it is a natural leap to consider other sources of external information, such as lifestyle, purchasing, household, social network, and environmental data, likely to be useful for making actuarial predictions [11, 24].

Second, the use of credit and other scoring models has served as an early example of a widening domain for predictive models in insurance. It is certainly natural for actuaries to employ modern analytical and predictive modeling techniques to arrive at better solutions to traditional actuarial problems such as estimating mortality, setting loss reserves, and establishing classification ratemaking schemes. But actuaries and other insurance analytics are increasingly using predictive modeling techniques to improve business processes that traditionally have been largely in the purview of human experts.

For example, the classification ratemaking paradigm for pricing insurance is of limited applicability for the pricing of commercial insurance policies. Commercial insurance pricing has traditionally been driven more by underwriting judgment than by actuarial data analysis. This is because commercial policies are few in number relative to personal insurance policies, are more heterogeneous, and are described by fewer straightforward rating dimensions. Here, the scoring model paradigm is especially useful. In recent years it has become common for scoring models containing a large number of commercial credit and non-credit variables to ground the underwriting and pricing process more in actuarial analysis of data, and less in the vagaries of expert judgment. To be sure, expert underwriters remain integral to the process, but scoring models replace the blunt instrument of table- and judgment-driven credits and debits with the precision tool of modeled conditional expectations.

Similarly, insurers have begun to turn to predictive models for scientific guidance of expert decisions in areas such as claims management, fraud detection, premium audit, target marketing, cross-selling, and agency recruiting and placement. In short, the modern paradigm of predictive modeling has made possible a broadening, as well as a deepening, of actuarial work.

As in actuarial science, so in the larger worlds of business, education, medicine, sports, and entertainment. Predictive modeling techniques have been effective in a strikingly diverse array of applications such as:

- Predicting criminal recidivism [12] Making psychological diagnoses [12]
- Helping emergency room physicians more effectively triage patients [13] Selecting players for professional sports teams [14]
- Forecasting the auction price of Bordeaux wine vintages [15]
- Estimating the walk-away "pain points" of gamblers at Las Vegas casinos to guide casino personnel who intervene with free meal coupons [15]
- Forecasting the box office returns of Hollywood movies [16]

A common theme runs through both these and the above insurance applications of predictive modeling. Namely, in each case predictive models have been effective in domains traditionally thought to be in the sole purview of human experts. Such findings are often met with surprise and even disbelief.

Psychologists, emergency room physicians, wine critics, baseball scouts, and indeed insurance underwriters are often and understandably surprised at the seemingly uncanny power of predictive models to outperform unaided expert judgment. Nevertheless, substantial academic research, predating the recent enthusiasm for business analytics by many decades, underpins these findings. Paul Meehl, the seminal figure in the study of statistical versus clinical prediction, summed up his life's work thus [17]:

> There is no controversy in social science which shows such a large body of quantitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing over 100 investigations, predicting everything from the outcome of football games to the diagnosis of liver disease, and when you can hardly come up with half a dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion.

Certainly not all applications of predictive modeling have a "clinical versus actuarial judgment" character [18]. For example, amazon.com and netflix.com make book and movie recommendations without any human intervention [25]. Similarly, the elasticity-optimized pricing of personal auto insurance policies can be completely automated (barring regulatory restrictions) through the use of statistical algorithms. Applications such as these are clearly in the domain of machine, rather than human, learning. However, when seeking out ways to improve business processes, it is important to be cognizant of the often surprising ability of predictive models to improve judgment-driven decision-making.

CURRENT STATE OF LIFE INSURANCE PREDICTIVE MODELING

While life insurers are noted among the early users of statistics and data analysis, they are absent from the above list of businesses where statistical algorithms have been used to improve expert-driven decisions processes. Still, early applications of predictive modeling in life insurance are beginning to bear fruit, and we foresee a robust future in the industry [19].

Life insurance buffers society from the full effects of our uncertain mortality. Firms compete with each other in part based on their ability to replace that uncertainty with (in aggregate) remarkably accurate estimates of life expectancy. Years of fine-tuning these estimates have resulted in actuarial tables that mirror aggregate insured population mortality, while underwriting techniques assess the relative risk of an individual. These methods produce relatively reliable risk selection, and as a result have been accepted in broadly similar fashion across the industry. Nonetheless, standard life insurance underwriting techniques are still quite costly and time consuming. A life insurer will typically spend approximately one month and several hundred dollars underwriting each applicant.[1]

Many marginal improvements to the underwriting process have taken hold: simplified applications for smaller face amounts, refinement of underwriting requirements based upon protective value studies, and streamlined data processing via automated software packages are all examples. However, the examples in the previous section suggest that property-casualty insurers have gone farther in developing analytics-based approaches to underwriting that make better use of available information to yield more accurate, consistent, and efficient decision-making. Based on our experience, life insurance underwriting is also ripe for this revolution in business intelligence and predictive analytics. Perhaps motivated by the success of analytics in other industries, life insurers are now beginning to explore the possibilities.[2]

Despite our enthusiasm, we recognize that life underwriting presents its own unique set of modeling challenges which have made it a less obvious candidate for predictive analytics. To illustrate these challenges it is useful to compare auto underwriting, where predictive modeling has achieved remarkable success, with life underwriting, where modeling is a recent entry. Imagine everything an insurer could learn about a prospective customer: age, type of car, accident history, credit history, geographic location, personal and family medical history, behavioral risk factors, and so on. A predictive model provides a mapping of all these factors combine onto the expected cost of insuring the customer. Producing this map has several prerequisites:

- A clearly defined target variable, i.e., what the model is trying to predict
- The availability of a suitably rich data set, in which at least some predictive variables correlated with the target can be identified
- A large number of observations upon which to build the model, allowing the abiding relationships to surface and be separated from random noise
- An application by which model results are translated into business actions

While these requirements are satisfied with relative ease in our auto insurance example, life insurers may struggle with several of them.

	Auto Insurer	Life Insurer
Target Variable	Claims over six-month contract	Mortality experience over life of product (10, 20+ years)
Modeling Data	Underwriting requirements supplemented by third-party data	Underwriting requirements supplemented by third-party data
Frequency of Loss	Approximately 10 percent of drivers make claims annually	Typically, fewer than 1 first year death per 1,000 new policies issued
Business Action	Underwriting Decision	Underwriting Decision

Statisticians in either domain can use underwriting requirements, which are selected based upon their association with insurance risk, supplement them with additional external data sources, and develop predictive models that will inform their underwriting decisions. However, the target variable and volume of data required for life insurance models raise practical concerns.

For the auto insurer, the amount of insurance loss over the six-month contract is an obvious candidate for a model's target variable. But because most life insurance is sold through long duration contracts, the analogous target variable is mortality experience over a period of 10, 20, or often many more years. Because the contribution of a given risk factor to mortality may change over time, it is insufficient to analyze mortality experience over a short time horizon. Further, auto insurers can correct underwriting mistakes through rate increases in subsequent policy renewals, whereas life insurers must price appropriately from the outset.

We recognize that life underwriting presents its own unique set of modeling challenges.

The low frequency of life insurance claims (which is good news in all other respects) also presents a challenge to modelers seeking to break ground in the industry. Modeling statistically significant variation in either auto claims or mortality requires a large sample of loss events. But whereas approximately 10 percent of drivers will make a claim in a given year, providing an ample data set, life insurers can typically expect less than one death in the first year of every 1,000 policies issued.[3] Auto insurers can therefore build robust models using loss data from the most recent years of experience, while life insurers will most likely find the data afforded by a similar time frame insufficient for modeling mortality.

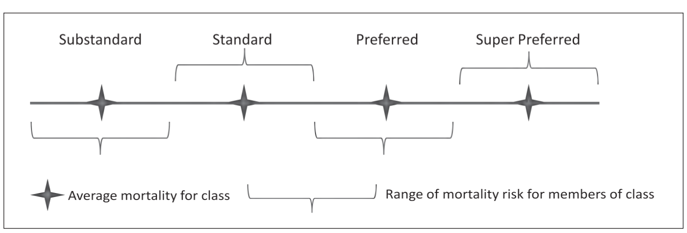
The low frequency of death and importance of monitoring mortality experience over time leaves statisticians looking for life insurance modeling data that spans many (possibly 20) years. Ideally this would be a minor impediment, but in practice, accessing usable historical data in the life insurance industry is often a significant challenge. Even today, not all life insurers capture underwriting data in an electronic, machine-readable format. Many of those that do have such data only implemented the process in recent years. Even when underwriting data capture has been in place for years, the contents of the older data (i.e., which requirements were ordered) may be very different from the data gathered for current applicants.

These challenges do not preclude the possibility of using predictive modeling to produce refined estimates of mortality. However, in the short term they have motivated a small, but growing number of insurers to begin working with a closely related yet more immediately feasible modeling target: the underwriting decision on a newly issued policy. Modeling underwriting decisions rather than mortality offers the crucial advantage that underwriting decisions provide informative short term feedback in high volumes. Virtually every application received by a life insurer will have an underwriting decision rendered within several months. Further, based upon both historical insurer experience and medical expertise, the underwriting process is designed to gather all cost-effective information available about an applicant's risk and translate it into a best estimate of future expected mortality. Therefore, using the underwriting decision as the target variable addresses both key concerns that hinder mortality- predicting models.

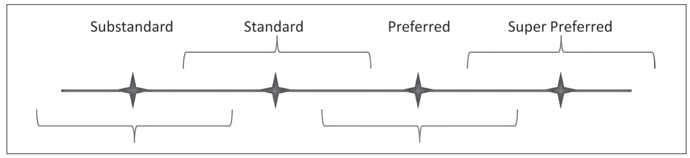
Of course, underwriting decisions are imperfect proxies for future mortality. First, life underwriting is subject to the idiosyncrasies, inconsistencies, and psychological biases of human decision-making.

Indeed this is a major motivation for bringing predictive models to bear in this domain. But do these idiosyncrasies and inconsistencies invalidate underwriting decisions as a candidate target variable? No. To the extent that individual underwriters' decisions are independent of one another and are not affected by common biases, their individual shortcomings tend to "diversify away." A famous example illustrates this concept. When Francis Galton analyzed 787 guesses of the weight of an ox from a contest at a county fair, he found that the errors of the individual guesses essentially offset one another, and their average came within one pound of the true weight of the ox. This illustrates how regression and other types of predictive models provide a powerful tool for separating "signal" from "noise."

In fact, the Galton example is quite similar to how life insurers manage mortality. Although individual mortality risk in fact falls along a continuum, insurers group policyholders into discrete underwriting classes and treat each member as if they are of average risk for that class. When the risks are segmented sufficiently, insurers are able to adequately price for the aggregate mortality risk of each class.



However, to avoid anti-selection and maintain the integrity of the risk pools insurers must segment risks into classes that are homogenous. While the "noise" in underwriting offers may diversify, these offers are accepted or rejected by applicants strategically. On average, applicants who have knowledge of their own health statuses will be more likely to accept offers that are in their favor, and reject those that are disadvantageous. For example, in the figure below an applicant at the upper range of the standard class may qualify for preferred with another insurer, thus leaving the risk profile of the original standard class worse than expected.



Therefore, anything that widens the range of mortality risks in each class, and thus blurs the lines between them, poses a threat to a life insurer. In addition to the inconsistency of human decision making, global bias resulting from company-wide underwriting guidelines that may not perfectly represent expected mortality can also contribute to this potential problem.

While modeling underwriting decisions may ultimately become a step along the path towards modeling mortality directly, we do believe today it is a pragmatic approach that provides the maximal return on modeling investment today. Specifically, utilizing underwriting decisions as the target variable is advantageous because they are in generous supply, contain a great deal of information and expert judgment, and do not require long "development" periods as do insurance claims. At the same time they contain diversifiable "noise" that can be dampened through the use of predictive modeling. Although building models for mortality and improving risk segmentation remain future objectives, utilizing predictive models based upon historical underwriting decisions represents a significant improvement on current practice, and is a practical alternative in the common scenario where mortality data is not available in sufficient quantities for modeling.

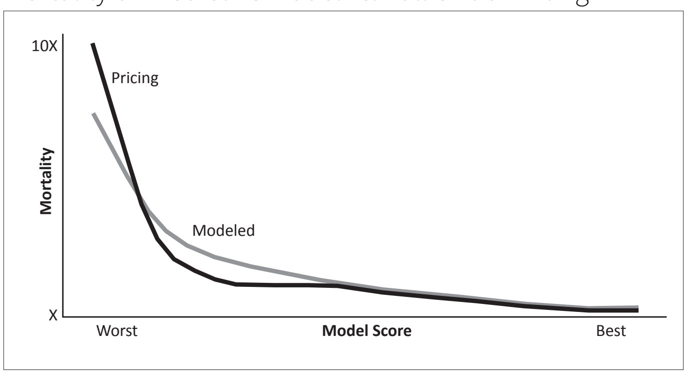
BUSINESS APPLICATION THAT CAN HELP DELIVER A COMPETITIVE ADVANTAGE

We will describe the technical aspects of underwriting predictive models in some detail in a subsequent section. While that discussion may beguile certain members of the audience (the authors included), others will be more interested in understanding how predictive modeling can deliver a competitive advantage to life insurers.

Life Underwriting

Unsurprisingly, one compelling application has been to leverage models that predict underwriting decisions directly within the underwriting process. As mentioned above, underwriting is a very costly and time consuming, but necessary, exercise for direct life insurance writers. Simply put, the underwriting process can be made faster, more economical, more efficient, and more consistent when a predictive model is used to analyze a limited set of underwriting requirements and inexpensive third-party marketing data sources (both described below) to provide an early glimpse of the likely underwriting result. As illustrated in Figure 1, the underwriting predictive models that Deloitte has helped insurers develop have been able to essentially match the expected mortality for many applicants. These insurers are beginning to leverage model results to issue many of their policies in just several days, thus foregoing the more costly, time consuming, and invasive underwriting requirements.

Figure 1
Mortality of Predictive Model vs. Full Underwriting



Risks which had been underwritten by the insurer and kept in a holdout sample were rank-ordered by model score and divided into equal-sized deciles. Modeled mortality is computed by taking a weighted average of the insurer's mortality estimates for each actual underwriting class in proportion to their representation within each decile. Pricing mortality represents the fully underwriting pricing mortality assumptions.

Issuing more policies with fewer requirements may initially seem like a radical change in underwriting practices, but we think of it as an expansion of a protective value study. Just as insurers currently must judge when to begin ordering lab tests, medical exams records, and EKGs, the models are designed to identify which applicant profiles do and do not justify the cost of these additional requirements. Based on the results of the models we've helped insurers build thus far, the additional insight they provide has allowed these insurers to significantly change the bar on when additional tests are likely to reveal latent risk factors. As indicated by the quality of fit between the model mortality and pricing assumptions, these models have been able to identify approximately 30 percent to 50 percent of the applicants that can be issued policies through a streamlined process, and thus avoid the traditional requirements.

With impressive frequency, the underwriting decision recommended by these models matched the decision produced through full underwriting. For cases when they disagree, however, we offer two possibilities: 1) the models do not have access to information contained in the more expensive requirements which may provide reason to change the decision, or 2) models are not subject to biases or bounded cognition in the same way that underwriters, who do not always act with perfect consistency or optimally weigh disparate pieces of evidence, are. The latter possibility comports with Paul Meehl's and his colleagues' studies of the superiority of statistical over clinical decision making, and is further motivation for augmenting human decision-making processes with algorithmic support.

In our analyses of discrepancies between models and underwriting decisions we did encounter cases where additional underwriting inputs provided valuable results, but they were rivaled by instances of underwriting inconsistency. When implementing a model, business rules are used to capitalize upon the model's ability to smooth inconsistency, and channel cases where requirements are likely to be of value to the traditional underwriting process. Thus, our experience therefore suggests that insurance underwriting can be added to the Meehl school's long list of domains where decision-making can be materially improved through the use of models.

These results point to potentially significant cost savings for life insurers. Based on a typical company's volume, the annual savings from reduced requirements and processing time are in the millions, easily justifying the cost of model development. Table 1 shows a rough example of the potential annual savings for a representative life insurer. It lists standard underwriting requirements and roughly typical costs and frequencies with which they would be ordered in both a traditional and a model-enhanced underwriting process. It then draws a comparison between the costs of underwriting using both methods.

Table 1
Illustrative Underwriting Savings from Predictive Model

	Requirement Cost	Requirement Utilization	
		Traditional Underwriting	Predictive Model
Paramedical Exam	$55	50%	0%
Oral Fluids Analysis	$25	20%	0%
Blood and Urine Analysis	$55	70%	0%
MVR Report	$6	70%	75%
Attending Physician Statement	$100	20%	0%
Medical Exam	$120	20%	0%
EKG	$75	10%	0%
Stress Test	$450	1%	0%
Third-Party Data	$0.50	0%	100%
Total Cost Per Applicant		**$130**	**$5**
Savings Per Applicant		**$125**	
Annual Applications Received		**50,000**	
Annual Savings (over 30% to 50% of applications)		**$2 to $3 million**	

In addition to hard dollars saved, using a predictive model in underwriting can generate opportunities for meaningful improvements in efficiency and productivity. For example, predictive modeling can shorten and reduce the invasiveness of the underwriting. The time and expense required to underwrite an application for life insurance and make an offer is an investment in ensuring that risks are engaged at an appropriate price. However, the effort associated with the underwriting process can be considered a deterrent to purchasing life insurance. Resources spent while developing a lead, submitting an application, and underwriting a customer who does not ultimately purchase a policy are wasted from the perspective of both the producer and home office. The longer that process lasts, and the more tests the applicant must endure, the more opportunity the applicant has to become frustrated and abort the purchase entirely, or receive an offer from a competitor. Further, complications with the underwriting process also provide a disincentive for an independent producer to bring an applicant to a given insurer. Enhancing underwriting efficiency with a model can potentially help life insurers generate more applications, and place a higher fraction of those they do receive. In addition, the underwriting staff, which is becoming an increasingly scarce resource,[4] will be better able to handle larger volumes as more routine work is being completed by the model.

We should emphasize that we do not propose predictive models as replacements for underwriters. Underwriters make indispensable contributions, most notably for applicants where medical tests are likely to reveal risk factors requiring careful consideration. Ideally, models could be used to identify the higher risk applicants early in the underwriting process, streamline the experience for more straightforward risks, and thus free up the underwriter's time for analysis of the complex risks. In addition, underwriters can and should provide insight during the construction, evaluation, and future refinements of predictive models. This is an oft overlooked but significant point. Particularly in complex domains such as insurance, superior models result when the analyst works in collaboration with the experts for whom the models are intended.

How exactly does the process work? The rough sequence is that the insurer receives an application, then a predictive model score is calculated, then a policy is either offered or sent through traditional underwriting. In more detail, the predictive model is typically used not to make the underwriting decisions, but rather to triage applications and suggest whether additional requirements are needed before making an offer. To that end, the model takes in information from any source that is available in near-real time for a given applicant. This can include third-party marketing data and more traditional underwriting data such as the application/tele-interview, MIB, MVR, and electronic prescription database records. For most insurers, this data can be obtained within two days of receiving the application.[5]

We should point out one key change some insurers must endure. It is essential that producers do not order traditional requirements at the time an application is taken. If all requirements are ordered immediately at the application, eliminating them based upon model results is impossible. For some insurers, this is a major process change for the producer group.

After loading the necessary data for model inputs, the model algorithm runs and produces a score for the application. From here, several approaches can lead to an underwriting decision. One central issue insurers may wrestle with is how to use the model output when justifying an adverse action (i.e., not offering an individual applicant the lowest premium rate). Due to regulatory requirements and industry conventions, it is customary to explain to applicants and producers the specific reasons in cases where the best rates are not offered. It is possible to fashion a reason message algorithm that "decomposes" the model score into a set of intuitively meaningful messages that convey the various high-level factors pushing an individual score in a positive or negative direction. There is considerable latitude in the details of the reason message algorithm, as well as the wording of the messages themselves.

While allowing the model algorithm to place applicants in lower underwriting classes while delivering reason codes is a viable, given the novelty of using predictive modeling in underwriting, the approach life insurers have been most comfortable with thus far is underwriting triage. That is, allowing the model to judge which cases require further underwriting tests and analysis, and which can be issued immediately. From a business application perspective, the central model implementation question then becomes: what model score qualifies an applicant for the best underwriting class that would otherwise be available based upon existing underwriting guidelines? The information contained in the application and initial requirements will set a ceiling upon the best class available for that policy. For example, let us assume an insurer has set an underwriting criterion that says children of parents with heart disease cannot qualify for super preferred rates. Then for applicants that disclose parents with this condition on the application, a model can recommend an offer at preferred rates without taking the decisive step in the disqualification from super preferred.

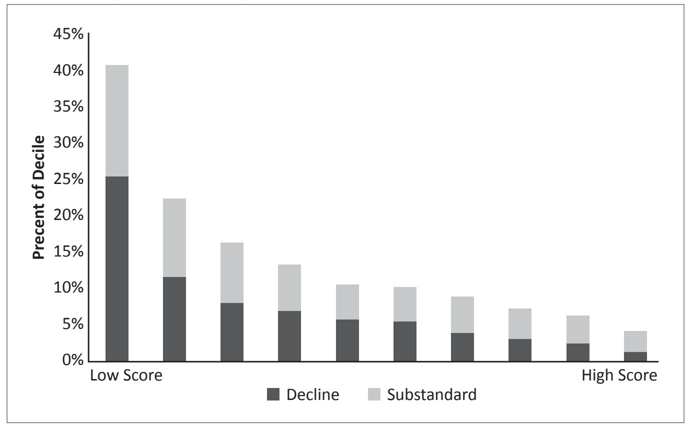
That is, the role of the model is to determine whether an applicant's risk score is similar enough to other applicants who were offered preferred after full underwriting. If so, the insurer can offer preferred to this applicant knowing the chance that additional requirements will reveal grounds for a further downgrade (the protective value) will be too small to justify their cost. If the applicant's risk score is not comparable to other preferred applicants, the insurer can continue with the traditional underwriting.

Marketing

In addition to making the underwriting process more efficient, modeling underwriting decisions can be of assistance in selling life insurance by identifying potential customers who are more likely to qualify for life insurance products. Marketing expenses are significant portions of life insurance company budgets, and utilizing them efficiently is a key operational strategy. For example, a company may have a pool of potential customers, but know little about their health risks at the individual level. Spreading the marketing efforts evenly over the pool will yield applicants with average health. However, this company could likely increase sales by focusing marketing resources on the most qualified customers.

The models supporting underwriting decisions that we have discussed thus far leverage both third-party marketing data and a limited set of traditional underwriting requirements. Alternatively, we can build predictive models using only the marketing data. While these marketing models do not deliver the same predictive power as those that utilize traditional underwriting data, they still segment risks well enough to inform direct marketing campaigns. Scoring the entire marketing pool and employing a targeted approach should help reduce the dollars spent marketing to those who will later be declined or less likely to accept an expensive offer, and result in an applicant pool that contains more healthy lives.

Figure 2
Marketing Model Segmentation



Like Figure 1, risks which had been underwritten by the insurer and kept in a holdout sample were rank-ordered by model score (using third-party data only) and divided into equal-sized deciles. However, this graph shows fractions of those deciles which contain declined or substandard applicants.

In addition to general target marketing efforts, models of underwriting decisions can also serve more specific sales campaigns. For example, multiline insurers, or even broader financial institutions often attempt to increase sales by cross-marketing life products to existing customers. However, they run the risk of alienating a current customer if the post-underwriting offer is worse than what the marketing suggests. Instead of selling an additional product, the company may be at risk of losing the customer. In dealing with this challenge, predictive modeling can be used to conduct an initial review of the customer pool and assist in determining which existing customers should receive offers for life insurance.

Predictive modeling can also aid in targeting specific products to the markets for which they were designed. For example, a given company may sell a product with preferred rates that are competitive, but standard rates that are less attractive. Other products may contain incentives for the insured to maintain healthy lifestyle. To whom might these products appeal? A person who the model indicates is currently living a healthy lifestyle is a prime target for such marketing programs.

In-Force Management

It is well known that the effects of underwriting wear off over time. Lives that were initially healthy may have degraded, and people who appeared to be poor risks initially may have improved. Products are priced to anticipate a reversion to mean health risk, but considerable variation in the health status of in-force policyholders will both remain and be unobservable without new underwriting. While full underwriting is cost prohibitive in these situations, a predictive model could be an inexpensive and transparent alternative. Scoring the in-force block could provide more insight to emerging mortality experience, inform changes to nonguaranteed policy features, help insurers know where to focus efforts to retain policyholders, and guide both direct writers and reinsurers in block transactions.

Additional Predictive Model Applications

We have focused our discussion on modeling health risk for life insurers because it is arguably the latest advancement, but there are many other areas of uncertainty for life insurers where a predictive model could reveal valuable information. We will present several potential applications in brief.

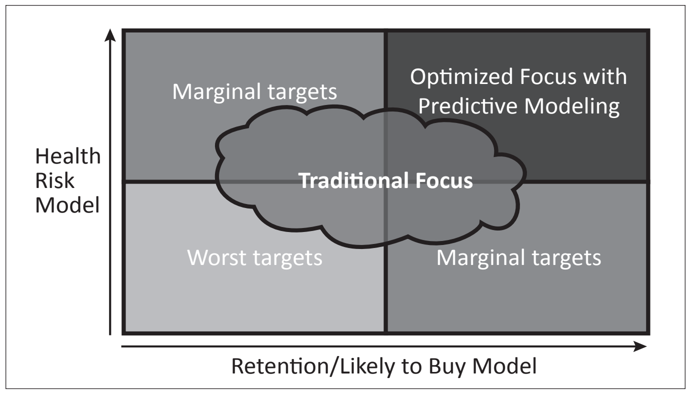
Analogous to models used to market consumer products, predictive algorithms can also estimate how interested a potential customer would be in purchasing a product from a life insurance company.

Insurance customers are often relatively affluent, or have recently undergone life-changing events such as getting married, having children, or purchasing a house. All of these traits and events (among other characteristics) can be identified in the marketing data. More specifically, a predictive model can be built to identify which characteristics are most highly correlated with the purchase of life insurance.

Again, scoring a direct marketing database can help a life insurer determine where to focus limited resources for marketing and sales.

We have discussed retention in terms of which customers an insurer would most like to keep, but equally important is which customers are most likely to leave. As many of the same life event and lifestyle indicators in the marketing data communicate when a person is likely to purchase a product, they also hint when a person is likely to surrender a product. In addition to third-party data, insurers also can see indicators of impending surrenders in transactional data such as how policyholders are paying premiums (automatic bank debits vs. having to physically write each year, or month), whether a policyholder is taking policy loans, whether they are calling the home office asking for cash values, account balances, and in-force illustrations, etc. Since neither producers nor the home office can give complete attention to each policyholder, a predictive model can sort these different indicators and help prioritize where to focus policy-saving effort.

Predictive modeling becomes even more powerful when models are used in combination. Not only can they answer who is most likely to purchase or surrender, but they can simultaneously identify the customers the company would most like to insure. Layering the underwriting health risk model on top of either the purchase or surrender models will tell the insurer which quadrant of the population will likely yield the highest return.



A final application we will mention is workforce analytics. Becoming a successful life insurance agent is notoriously difficult. The home office spends significant resources recruiting and training these agents, and the high turnover rate is a considerable drain. Predictive models can be used to help improve the efficiency of agent recruiting by scoring applicants on the basis of how similar their profile is to that of a company's existing successful field force. Such a tool can help prioritize which applicants to pursue.

When considering all the potential applications for predictive modeling in life insurance, it becomes apparent that analytics is truly an enterprise capability rather than a niche solution. After an insurer begins with one application that proves successful, the next application follows more easily than the first. Expertise, data, and infrastructure can be leveraged throughout the organization, but more importantly, decision makers come to realize and respect the power of predictive modeling.

As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see *www.deloitte.com/us/about* for a detailed description of the legal structure of Deloitte LLP and its subsidiaries.

Chris Stehno is managing director at Deloitte Consulting LLP. He can be reached at *cstehno@deloitte.com*.

Jim Guszcza is the US chief data scientist of Deloitte Consulting, and a member of Deloitte's Advanced Analytics and Modeling practice. He can be reached *at jguszcza@deloitte.com*.

ENDNOTES

1 According to the Deloitte 2008 LIONS benchmarking study of 15 life insurers, the median service time to issue a new policy ranges between 30 and 35 days for policies with face amounts between $100k to $5 million, and the average cost of requirements (excluding underwriter time) is $130 per applicant.

2 As reported in an SOA sponsored 2009 study, "Automated Life Underwriting," only 1 percent of North American life insurers surveyed are currently utilizing predictive modeling in their underwriting process.

3 This is an estimate based upon industry mortality tables. Mortality experience varies across companies with insured population demographics. In the 2001 CSO table, the first-year select, weighted average mortality rate (across gender and smoker status) first exceeds 1 death per thousand at age 45.

4 According to the Bureau of Labor Statistics 2010-2011 Occupational Outlook Handbook, despite reduced employment due to increased automation, the job outlook of insurance underwriters is classified as "good" because "the need to replace workers who retire or transfer to another occupation will create many job openings."

5 Receiving the application is defined as when all application questions have been answered and/or the tele- interview has been conducted. If applicable, this includes the medical supplement portion of the application.

REFERENCES

1 Davenport, T. H., Harris, J. G., *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, (2006).

2 Guszcza, J., "Analyzing Analytics*", Contingencies,* American Academy of Actuaries, July- August, (2008).

3 McCullagh, P., Nelder, J. A., *Generalized Liner Models*, 2nd Edition, Chapman and Hall, (1989).

4 Brockman, M. J., Wright, T. S., "Statistical Motor Rating: Making Effective Use of Your Data," *Journal of the Institute of Actuaries*, Vol. 119, Part III, (1992).

5 Mildenhall, S.J., "A Systematic Relationship between Minimum Bias and Generalized Linear Models," *Proceedings of Casualty Actuarial Society*, Vol. LXXXVI, Casualty Actuarial Society, (1999).

6 Bailey, R. A., Simon, L. J., "An Actuarial Note on the Credibility of a Single Private Passenger Car", *Proceedings of Casualty Actuarial Society*, Vol. LVII, Casualty Actuarial Society, (1959).

7 Bulhmann, H., "Experience Rating and Credibility", *ASTIN Bulletin IV*, Part III, (1967).

8 "Insurance Scoring in Private Passenger Automobile Insurance – Breaking the Silence," *Conning Report,* Conning, (2001).

9 *Report on the Use of Credit History for Personal Line of Insurance*, American Academy of Actuaries, (2002).

10 "The Top 10 Casualty Actuarial Stories of 2009," *Actuarial Review*, Vol. 37, No. 1, Casualty Actuarial Society, (2010).

11 Guszcza, J., Wu, C. P., "Mining the Most of Credit and Non-Credit Data", *Contingencies,* American Academy of Actuaries, March-April, (2003).

12 Meehl, P., *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, University of Minnesota Press, (1954).

13 Gladwell, M., *Blink: The Power of Thinking without Thinking*, Little Brown and Co, (2005).

14 Lewis, M., *Moneyball: the Art of Winning an Unfair Game,* W. W. Norton & Company, (2003).

15 Ayres, I., *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart,* Bantam Books, (2007).

16 Gladwell, M., "The Formula", *The New Yorker*, (2006).

17 Meehl P., "Causes and Effects of My Disturbing Little Book," *Journal of Personality Assessment*, Vol. 50, (1986).

18 Dawes, R., Faust, D., Meehl, P., "Clinical Versus Actuarial Judgment," *Science,* 243: 1668-1674, (1989).

19 Guszcza, J., Batty, M., Kroll, A., Stehno, C., "Bring Predictive Models to Life," *Actuarial Software Now - A Supplement to Contingencies*, American Academy of Actuaries, Winter, (2009).

20 Hastie, T., Tibshirani, R., Friedman, J. H., *The Elements of Statistical Learning*, Springer, (2003).

21 Berry, M. J. A., Linoff, G. S., *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, John Wiley & Sons, (2003).

22 Breiman, L., Friedman, J., Olshen, R., Stone, C., *Classification and Regression Trees*, New York: Chapman and Hall, (1993).

23 Yan, J., Masud, M., Wu, C. P., "Staying Ahead of the Analytical Competitive Curve: Integrating the Broad Range Applications of Predictive Modeling in a Competitive Market Environment," *CAS E-Forum*, Winter, Casualty Actuarial Society, (2008).

24 Stehno, C., Johns, C., "You Are What You Eat: Using Consumer Data to Predict Health Risk," *Contingencies*, American Academy of Actuaries, Winter, (2006).

25 Yi Zhang , Jonathan Koren, Efficient bayesian hierarchical user modeling for recommendation system, Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, July 2007, Amsterdam, The Netherlands.