Article from

**The Actuary**

# THE USE OF PREDICTIVE ANALYTICS IN THE DEVELOPMENT OF EXPERIENCE STUDIES

Recently, predictive analytics has drawn a lot of attention in the North American life insurance industry. In fact, many life insurers are already applying predictive modeling techniques, and it appears there are several promising new areas for the life insurance business. We are also seeing predictive analytics emerging across health insurance, pensions, annuities and group insurance settings. This growing interest in the potential of predictive analytics follows extensive use over the past decade in the North American property and casualty (P&C) business.

This article will focus on the use of predictive analytics in one specific area: the development of experience studies. We will also explore the pros and cons of using predictive analytics versus traditional techniques, and the implications for setting assumptions and improving profitability. While we focus here on examples for life insurance and variable annuities, the methods, ideas and conclusions may be applied across other insurance types as well.

## A TRADITIONAL APPROACH FOR DEVELOPING EXPERIENCE STUDIES

Historically, the typical approach used to develop experience studies in the North American life industry—which we define as individual life fully underwritten mortality, but also generally comparable for assumption for other major underwriting types—is as follows:

**❶ |** Determine appropriate experience period—often five years (but could vary).
**❷ |** Gather applicable data for this period:
  ➔ Factors that impact mortality
  ➔ Exposures and deaths, based on count and face amount.
**❸ |** Determine if certain records should be excluded (e.g., conversions and substandard policies).
**❹ |** Determine an appropriate basis for expected mortality:
  ➔ This is often expressed as an industry table (e.g., 2008 VBT RR 100)
  ➔ Or possibly a company-specific table (e.g., Acme Life's 2012 mortality table).

**IMPROVING TECHNIQUES**

BY JOHN FENTON AND KENDRICK LOMBARDO

**⑤ |** Calculate actual results and display the actual-to-expected (A/E) ratios for the business.

**⑥ |** Determine appropriate splits for displaying the A/E results:
- ➔ Males vs. females
- ➔ Issue age groupings (e.g., decennial)
- ➔ Durations or durational groupings (e.g., quinquennial for select period, then ultimate period)
- ➔ Face amount bands (e.g., $0–$99K, $100–$249K, $250–$499K, $500–$999K, $1M+)
- ➔ Risk class (e.g., best preferred nonsmoker (NS), second best preferred NS, standard NS, smoker).

**⑦ |** Companies also often differentiate their experience by relevant underwriting eras (e.g., issues of 2011 and later, 2002 to 2010, and pre-2002), where there were significant differences in underwriting practice.

**⑧ |** Other splits are sometimes examined, such as product type and distribution channel. Results by calendar year are often shown—generally at an aggregate level.

**⑨ |** Companies typically summarize the overall A/E for main components of their business (e.g., best preferred NS = 70%), as well as for various splits of the business (males, best preferred NS, face amount of $500–$999K, durations 1–5).

**⑩ |** Ideally, companies undertake a more detailed study once a year, but then more frequently provide summary A/Es on an aggregate basis—typically quarterly.

**⑪ |** The results from the experience studies are then considered, along with comparable industry tables/assumptions, to set assumptions. The credibility of company data is a consideration here.

### PREDICTIVE ANALYTICS—A NEW APPROACH

With predictive analytics, many of the core components of the traditional approach are also used—seriatim data over the past $n$ years, and inclusion of factors that are believed to drive mortality. A company could choose to stop there and not expand its experience data; however, predictive analytics allows a company to leverage much more data. A predictive analytics approach allows the examination of the relationships between the variables that are being examined on an all-else-equal basis. This cannot be done with traditional techniques without distorting factors and results. A description of a typical predictive analytics project process is shown in the sidebar below.

# how-to

## KEY STEPS IN A PREDICTIVE ANALYTICS PROJECT

**PROJECT SCOPE:** Determine your target variable (that which you are trying to predict), block of business (i.e., products, riders, issue years, etc.), timelines, budget and resources. It is also helpful to assemble the core team involved in the project, including predictive analytics as well as product and functional experts from across the company. Having both technical and business experts is critical to having a successful project.

**DATA COLLECTION AND VALIDATION:** The most critical step in the process—without clean data a model can be worthless and/or totally misleading. Consider both the number of years of historical data and the breadth of variables. It is important not to limit the variable inclusion to only those you think may be predictive. The best models are those that reveal "aha!" moments through creative data inclusion.

**INITIAL FACTOR ANALYSIS:** Typically univariate analysis to determine the factors that appear to be driving the results. "Univariate" analysis means that you look at each variable on its own to see if it may be predictive of your target. Note: Variables that look promising during the univariate analysis could drop out in the next stage due to correlations (two or more variables that are highly correlated, e.g., attained age and duration).

**MODEL BUILDING:** This step involves selecting the form of the model (e.g., multiplicative vs. additive); determining the factors to be included using a range of tests, including statistical tests and business knowledge; and determining which interactions should be included and what simplifications (groups and curves) are appropriate.

**MODEL VALIDATION:** Use techniques such as comparing A/E values on hold-out samples to validate the model. Hold-out samples could either reflect the most recent data or a random subset of all data.

A sample outcome from using predictive analytics to perform an experience study on the same block of individual life fully underwritten business noted in the traditional analysis is shown below. Note that in order to keep it comparable with the traditional approach, we are expressing the results as a percentage of an industry table (2008 VBT RR 100). However, we could also have expressed these results as a percentage of a new base table that we created as part of the process.

### EXPERIENCE FOR CALENDAR YEARS 2009 TO 2013

Base table: M/F, NS/SM version of 2008 VBT RR 100
Predictive factors included in the model:

❶ | Gender
❷ | Issue age bands
❸ | Policy durational groupings
❹ | Risk classes
   ⮑ Non-substandard classes
❺ | Face amount bands
❻ | Product type
   ⮑ Variable universal life (VUL), universal life (UL), term, other life
❼ | Type of underwriting tests employed
   ⮑ Non-medical
   ⮑ Medical with bodily fluids
   ⮑ Above with pharmacy check added
   ⮑ Above with attending physician's statement (APS) added
❽ | Distribution channel
   ⮑ Property and casualty (P&C) sales force
   ⮑ Career life sales force
   ⮑ Independent life agents
   ⮑ Financial advisers
❾ | Geo-demographic information
   ⮑ Based on ZIP/postal code

The results for each factor (or combination of factors) would generally be displayed as a multiplier to the base table. (Alternatively, they could also be expressed as an addition rather than a multiplier, although this is not typical.) The multipliers can take one of several forms: scalar, vector, matrix or via a formula (e.g., cubic).

Some of the above factors have historically not been used much in traditional experience studies. Here is more background on a couple of them:

### Type of underwriting test

Generally speaking, companies have not used the type of underwriting test as a factor in setting the mortality

**FINAL CALIBRATION:** This step includes refitting the model on all data, and taking into account any adjustments and final smoothing.

**IMPLEMENTATION:** This step involves taking the results of the project and incorporating them into the business purpose, often a model of the expected financial results (i.e., an underwriting program, a pricing model or valuation model). This could take one of several forms—a scoring engine might be developed for underwriting or another indicative risk selection process, or an enhanced calculation engine for other assumptions. Another consideration is how reports or results from the model will be used on an ongoing basis, as well as the planned update cycle for updating for new experience data.

assumption. Rather, this is reflected indirectly in the face amount band and possibly risk class (i.e., only possible for individuals aged 50 and up to obtain an average of $500K by undergoing extensive underwriting). We believe the type of underwriting test employed could offer further predictive value.
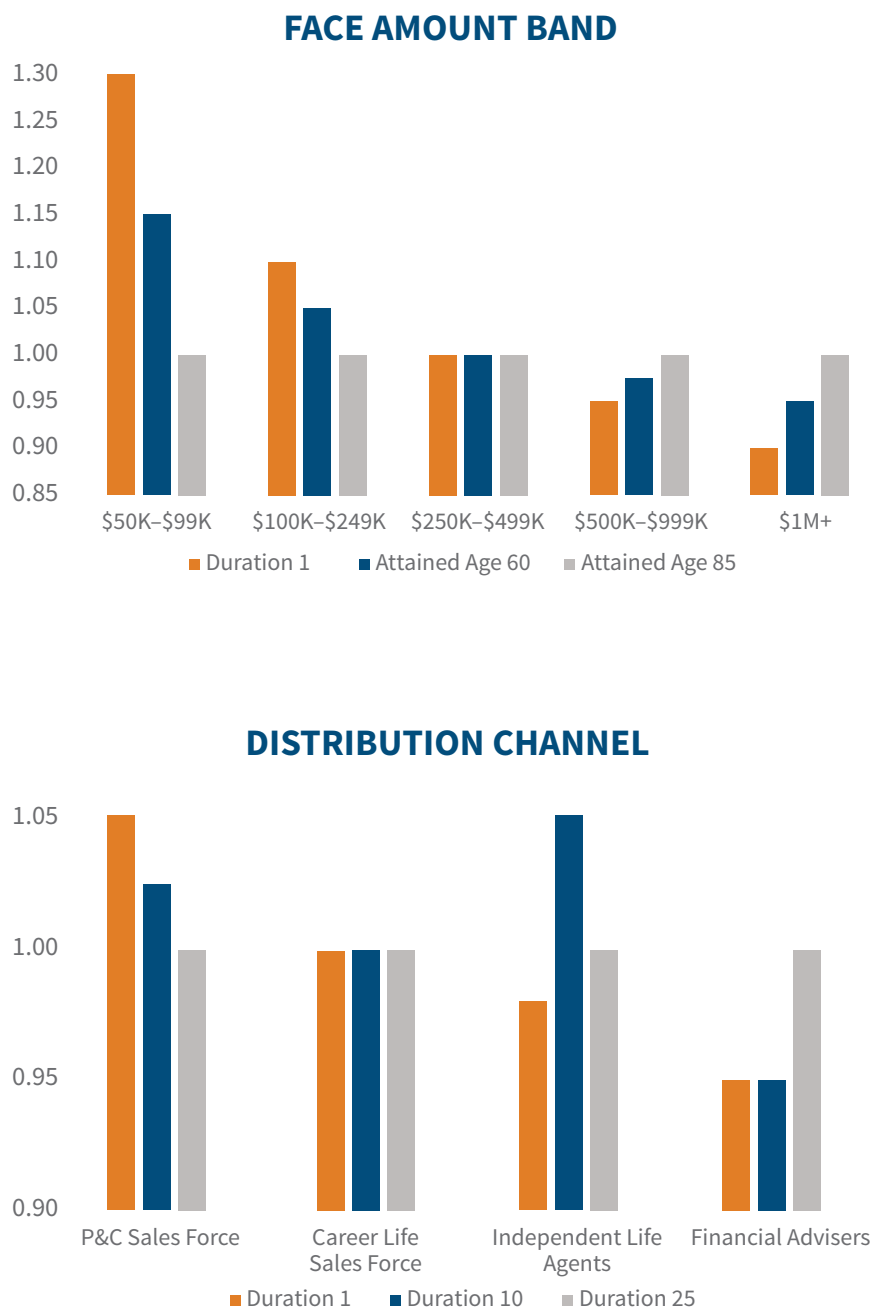
An extension of this would be to move beyond the type of test and examine the specific underwriting results (e.g., blood pressure of 140/85, LDL of 110, BMI of 31). However, availability of this underwriting data is often limited, which limits use of this technique. To the extent that it can be added, it would likely be a powerful predictor.

*Geo-demographic*

Other parts of the insurance industry have realized that where one lives can be a significant predictor of risk. This can be analyzed by accessing ZIP/postal code and associated geo-demographic information. The latter is information available from many sources, including the U.S. Census Bureau. It allows one to categorize key aspects of the socio-economic composition of a particular ZIP code (e.g., per capita income, percent with a bachelor's degree, homeownership, and residency in same house over past five years). The ZIP codes are mapped to the geo-demographic variables. Then each geo-demographic variable is grouped into a "sensible" number of levels that can be analyzed in the model. Advanced techniques are available to analyze residual effects.

Multipliers for some sample factors are provided as an example of how this might look. (See Figure 1.) These are not intended to be a full list and are shown as a scalar for simplicity, but more typically take another form.

**FIGURE 1:** MULTIPLIERS FOR SAMPLE FACTORS



**FACE AMOUNT BAND**

Legend: Duration 1, Attained Age 60, Attained Age 85

**DISTRIBUTION CHANNEL**

Legend: Duration 1, Duration 10, Duration 25

Hopefully, this provides a perspective on how experience studies could be undertaken for a block of individual life business using a predictive analytics approach. The sidebar below shows comparable results for a block of variable annuity (VA) business—studying policyholder behavior.
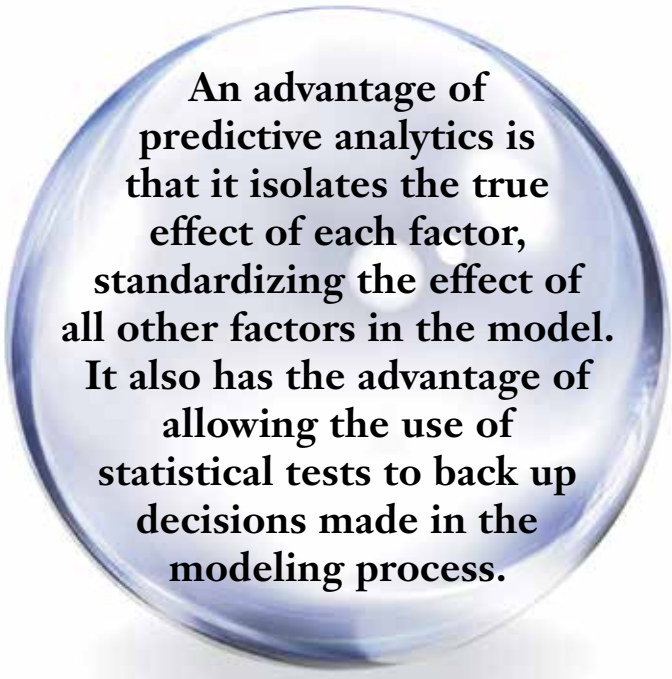
# findings

## THE IMPACT OF PREDICTIVE ANALYTICS ON VA EXPERIENCE STUDIES

VA designs have become increasingly complex, with faster evolutions, as companies try to respond to changing market conditions and increasingly sophisticated buyers and agents. We estimate that more than half of the top VA writers are using some type of predictive analytics techniques to project living benefit dynamic lapses. Predictive analytics techniques can also solve the dilemma of the interplay between base lapses and dynamic lapses. With traditional techniques, base lapses are required to estimate dynamic lapses, while dynamic lapses are required to estimate base lapses. With predictive analytics techniques, it is possible to solve for both simultaneously. Predictive analytics techniques have allowed companies to add increased sophistication to their models, with the necessary statistical rigor.

Predictive models have developed factors for the following:
- Base and dynamic lapses simultaneously
- Living and death benefits
- Rate/cap competitiveness
- Surrender changes and market value adjustment (MVA)
- Dynamic lapse sensitivity that varies by policy size
- Monthly lapse skew
- Factors for attained age, gender, tax status
- Benefit richness
- Interest-sensitive in-the-moneyness factors
- Distribution channel and product effects

### PROS AND CONS

There are, naturally, pros and cons for each approach. One advantage of the traditional approach is that it is widely used in the life insurance business and is commonly accepted. The processes used to support it are well-established; thus, it is relatively easy to produce new results. Also, management is presumably used to seeing the results and is able to determine how to apply the findings to undertake management action.

A primary advantage of using predictive analytics in developing experience studies is that it provides better insight into the interaction of various factors and allows for the better use of available data. For example, suppose one wanted to analyze the appropriate assumption for the male best preferred NS risk class (which represents 20 percent of the face amount in total) at face amounts of $1M+ (which represent 10 percent of the face amount in total). In this situation, it is likely a small portion of the business in total (the combination of male best preferred NS risk class at face amounts of $1M+ is likely higher than 2 percent due to the intersection of better risk classes at high face amounts, but still unlikely to exceed 4 to 5 percent) would be examined. This is also before considering other splits that one would like to examine, such as issue age band and durational grouping. Predictive analytics allows one to examine all of the data for male best preferred NS and for $1M+ face amount band, and determine what impact these factors have on experience.

> An advantage of predictive analytics is that it isolates the true effect of each factor, standardizing the effect of all other factors in the model. It also has the advantage of allowing the use of statistical tests to back up decisions made in the modeling process.

An advantage of predictive analytics is that it isolates the true effect of each factor, standardizing the effect of all other factors in the model. It also has the advantage of allowing the use of statistical tests to back up decisions made in the modeling process.

As such, use of predictive analytics allows one to introduce new factors (e.g., geo-demographic, type of underwriting) and evaluate their impacts without having to rely on traditional A/E results for increasingly smaller blocks of business—which would not be credible.

We should acknowledge that predictive analytics is a relatively recent introduction to the North American life insurance space, but it also requires specific expertise to implement. As such, there are dangers that data could be misinterpreted and models could be built with flaws. That said, any increase in sophistication brings risks that can be managed with the appropriate level of expertise and experience.

Under either the traditional or predictive analytics approach, there are some areas that cannot be fixed/addressed:

*Lack of data:* If experience data does not exist past attained age 85, it is difficult to solve the often-discussed issue: What is an appropriate mortality assumption at older ages? Predictive analytics enables the development of a trend line at older ages, including extrapolating past age 85, which arguably leads to better data than traditional analysis. However, it still requires judgment. Predictive analytics techniques can also make it easier to apply industry data to a company without significant data by adjusting the standard model based on the factors developed from the larger dataset.

*Not doing the study:* One of the primary areas where we see companies fall short is not doing an experience study, or not on a routine basis ("we don't have the resources right now"). One has to devote the time and resources to do the work. Correspondingly, an initial experience study using predictive analytics will likely take more time. This may be a consideration in undertaking a predictive analysis.

Another issue to consider is whether your models can support the more refined assumptions that come out of the predictive analytics approach. This is likely to be more of an issue for valuation models, as opposed to pricing models, which are smaller in size. Having the greatest factors in the world won't necessarily help if the more refined assumptions in the applicable model are not reflected.

## IMPLICATIONS FOR SETTING ASSUMPTIONS

A predictive analytics approach to experience studies leads to more refined assumptions, reflecting more granularity. The benefits of this are fairly evident: It allows one to better assess value—either value created from new business sales or value created from management of one's in-force block of business. A simple example that draws on a VA block of business can help explain this. The current lapse assumption varies by product type (i.e., share type) and duration. However, it does not distinguish between the following factors:

- ⮞ Age
- ⮞ Size of policy
- ⮞ Richness of living benefit feature.

The following chart shows the present value (PV) of illustrative profits at issue for selected cells (reflecting a new lapse formula by age is not shown directly below; rather, it is implicitly reflected in changes in the profit levels for the cells shown in Figure 2):

**FIGURE 2:** **PRESENT VALUE OF PROFITS AT ISSUE**

| PV of Profits | Revised Assumptions | | |
| --- | --- | --- | --- |
| | Benefit A | | Benefit B |
| | Small Policy | Large Policy | Small Policy |
| IA 60—No Wait | 0.30 | -0.10 | 0.15 |
| IA 60—10 Yr. Wait | 1.30 | 0.60 | 1.05 |
| IA 70—No Wait | 0.90 | 0.40 | 0.70 |
| IA 70—10 Yr. Wait | 1.70 | 1.00 | 1.45 |
| Total | 1.05 | 0.48 | 0.84 |

Thus, when the new assumption is reflected, there is a small increase in profitability from 0.63 percent to 0.73 percent. However, the key aspect is the management information embedded in the new profit pattern. Certain cells show significantly higher profitability than others. Based on this, management should take action—either to revise the benefit offerings at cells with adverse profitability (including possibly raising fees) or use other techniques to shift the mix of business (e.g., wholesaler incentives).

## THE CLEAR IMPLICATIONS OF PREDICTIVE ANALYTICS

The implications of using predictive analytics to undertake experience studies for setting assumptions are fairly clear.

- ⮞ The P&C industry has used predictive analytics for over a decade. Companies that don't use this face a serious disadvantage in the marketplace and the risk of anti-selection.
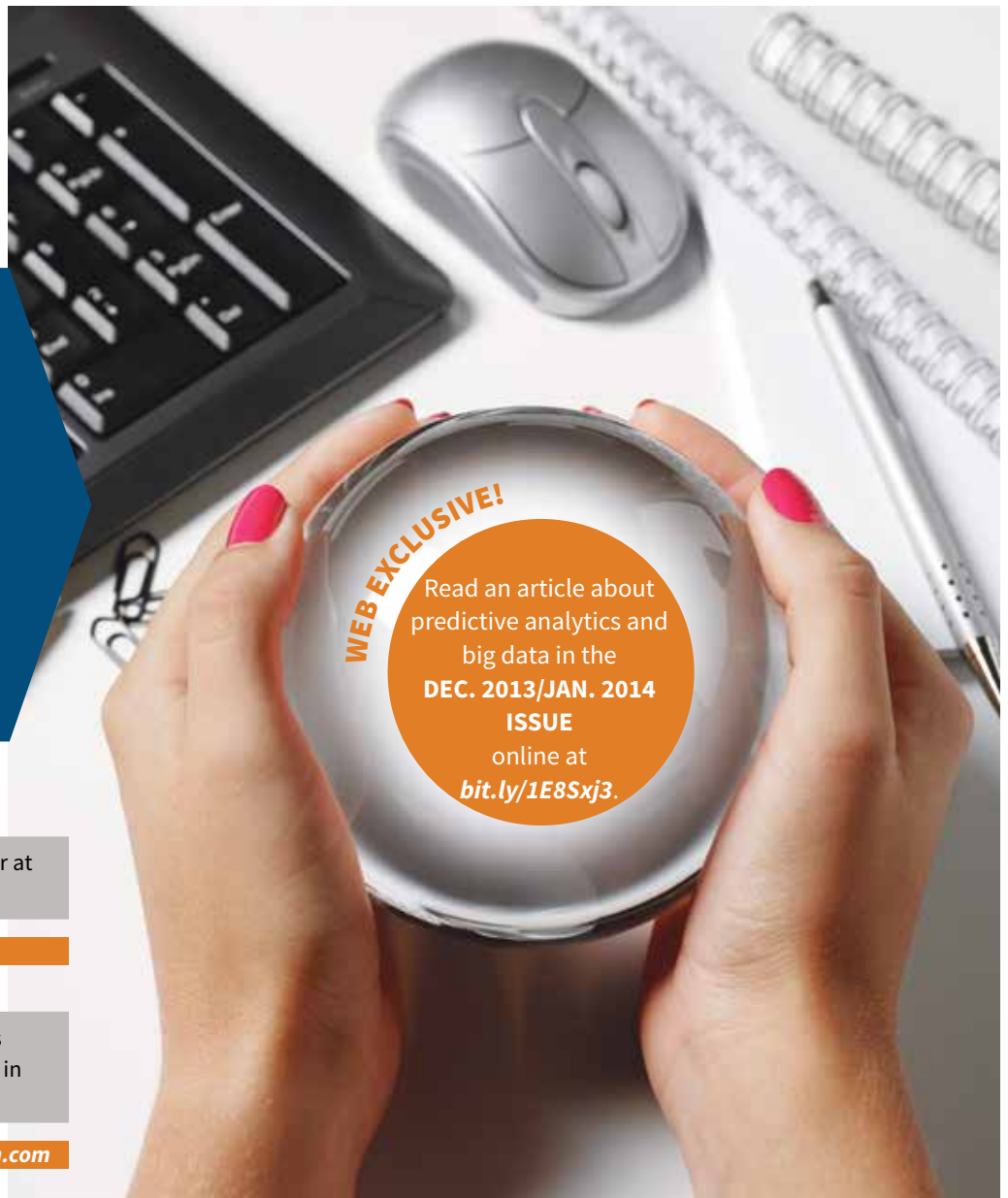
| | | Original Assumption |
|---|---|---|
| **Large Policy** | **Total** | **Total** |
| -0.10 | 0.06 | 0.00 |
| 0.70 | 0.91 | 0.80 |
| 0.45 | 0.61 | 0.50 |
| 1.10 | 1.31 | 1.20 |
| 0.54 | 0.73 | 0.63 |

**Predictive analytics can offer a new approach to experience studies that refine assumptions, are more granular, and ultimately create more clarity around the value of new and existing business.**

➡ Many companies have started to use predictive analytics techniques for experience studies and/or have also built big data teams to leverage predictive analytics across their organizations in other ways.

➡ Regulators are requesting increased sophistication in assumption setting, especially for principle-based reserves and capital techniques.

➡ Insurance company products are complex and evolve quickly. That makes it difficult to get a large dataset of homogeneous experience data. Predictive analytics can help leverage experience across a variety of designs and market conditions.

➡ There is more external data available than ever before, and companies are starting to explore how they can leverage it.

Traditional techniques can make it difficult to isolate the true effect of individual factors and to determine interactions or analyze the effect of changes in the mix of business (e.g., changes corresponding to changes in distribution). Predictive analytics, however, can offer a new approach to experience studies that refine assumptions, are more granular, and ultimately create more clarity around the value of new and existing business. ◼

**Companies that don't use predictive analytics could face a serious disadvantage in the marketplace.**

WEB EXCLUSIVE!
Read an article about predictive analytics and big data in the
**DEC. 2013/JAN. 2014 ISSUE**
online at
*bit.ly/1E8Sxj3*.

**John Fenton, FSA, MAAA,** is director at Towers Watson in Atlanta.

*John.Fenton@towerswatson.com*

**Kendrick Lombardo, FSA, MAAA,** is senior consultant at Towers Watson in Hartford, Connecticut.

*kendrick.lombardo@towerswatson.com*