Article from

Health Watch

December 2015
Issue 79

Producing Actionable Insights from Predictive Models Built Upon Condensed Electronic Medical Records

By Sheamus Kee Parkes

In April 2015, the research arm of the SOA published a set of papers on predictive analytics and how these techniques can be used in practice. One paper has been selected for publication in this issue of Health Watch; *however, readers are encouraged to check out the following link and read the rest of the publication. Please refer to this link for the entire publication: http://www.soa.org/News-and-Publications/Publications/Essays/2015-predictive-analytics.aspx.*

* * *

Predictive modeling often has two competing goals: accuracy and inference. In health care, risk scoring is used to make different groups more comparable and to explore drivers of costs. With care coordination specifically, patients need to be prioritized for intervention while also understanding why a given patient was prioritized. Care coordination can benefit from custom trained models that adapt to service patterns and include any novel sources of available information. These custom models can include industry-leading risk scores as inputs to retain their strengths and insights. One important novel input could be electronic medical records (EMR) data.

Predictive modeling with EMR is commonly associated with mining physicians' notes for nuanced opinions not found in the coarse diagnosis coding of medical claims. Although valuable, physician notes are not the only information in EMRs; other novel pieces of information include vitals measurements and lab results. Vitals information includes items such as height, weight, and blood pressure. Labs information includes results of panels such as lipid, metabolic, and blood counts. These too can provide a more nuanced view of a patient's health than demographics and claims alone. This article will recount the process of including labs and vitals information into a set of custom models built for care coordination efforts and then understanding the added value in accuracy and insights.

OBTAINING AND STANDARDIZING

The first hurdle in utilizing EMR information is obtaining it; it is often stored separately from claims data and under control of different staff or even a different organization. EMR table structure is commonly even less standardized than claims tables. Limiting to just vitals and labs makes the acquisition process easier. Once acquired, the labs and vitals information need similar, but not identical, processes to make them useful in predictive modeling.

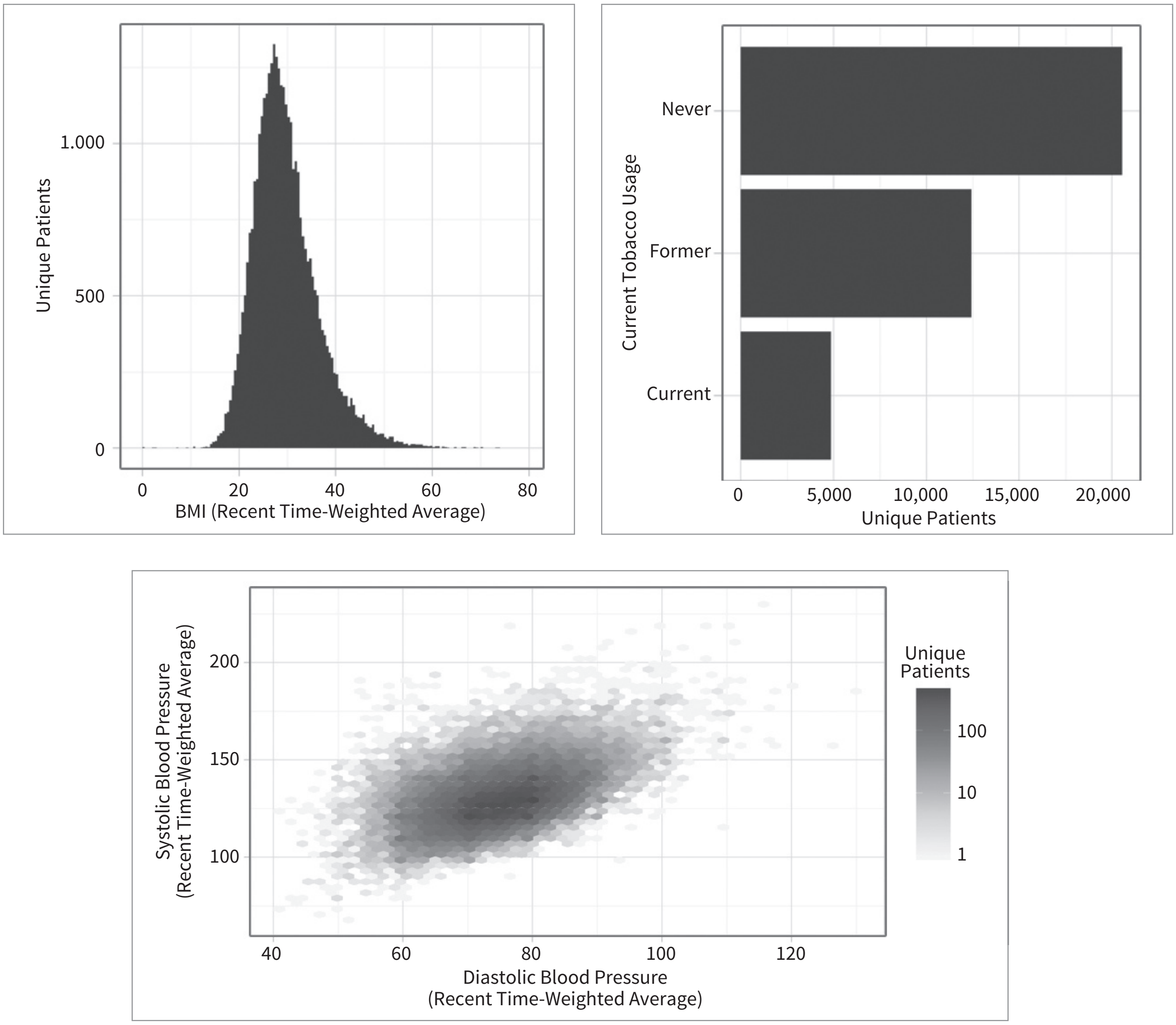
Labs and vitals both are needed on a timeline basis. Just having the most recent results for each patient would not be helpful unless pre-trained models were available that expected them as inputs. When training custom prospective models, a strong history of measurements is needed.

Labs and vitals are both subject to measurement and transcription errors. Although there is some clinical guidance available, concepts from robust statistics are invaluable in estimating useful bounds for outliers. Most items have generally symmetrical distributions of results.

While vitals data is collected more frequently than lab data, there are fewer types of information captured. Figure 1 shows the distribution of some key vitals information.

> Possibly more important, the EMR features provided new and potentially more actionable reasons for a given patient's predictions.

Figure 1
Distribution of EMR Vitals Information



Choosing among all these encoding possibilities can be somewhat of an art. However, it should be influenced by what learning algorithms will be applied.

Lab tests present additional hurdles. Results are collected from a variety of brick and mortar labs, and typically these entities do not report on a consistent basis. Most grievous is the lack of consistent naming of the item tested. For example, the following terms—BA%, BASOPHILS, Basophils %, and BASO%—all mean the same thing, which is separate from BA#, ABSOLUTE BASOFILS, and BASO (ABSOLUTE). A parsing library must be developed to standardize and categorize the labs data into consistent panel groups and individual items.

BUILDING FEATURE VECTORS

In health care, many analyses use patients as the units of observation. To perform analysis at a patient level, a useful feature vector needs to be built for each patient for each pertinent time period. When training custom models at least two time periods are needed: a historical training feature period for which future outcomes are known, as well as a current prediction feature period for which future outcomes are not yet known (but are of interest).

Within each feature period a given patient may have many measures of a given vital or lab, or none at all. There are many useful ways to collapse these sporadic time series. Simple possibilities would include taking the most recent value or a straight average of all recorded values. A slightly more refined approach would be a weighted average that gave more credit to recent values; this can strike a nice balance between freshness of information and measurement error reduction. There are seldom enough measurements per member to estimate a trend, but differences between first/last and minimum/maximum can be interesting, as can the simple count of the number of measurements of each item. Missing values are coded for those items a patient did not have measured at all.

Choosing among all these encoding possibilities can be somewhat of an art. However, it should be influenced by what learning algorithms will be applied. A reasonable choice of algorithm could be ensembled decision trees, primarily because they gracefully handle missing values, nonlinearities, and interactions while maintaining excellent performance. They can also utilize random feature sampling similar to that championed by Random Forests, so having modestly redundant features can be tolerated, as long as the included EMR features are not so plentiful that the more standard claims and eligibility features become diluted.

TRAINING MODELS AND ESTIMATING EFFECTS

Once the feature vectors are created, reasonable outcomes need to be chosen. Care coordination is often focused on avoiding the worst near-term outcomes, so useful outcomes can include the median and tail risk of total costs for the next six months.

Ensembled decision trees provide useful insights into what features are important. In this example, the claims-based features were still the most important, but the EMR features provided a small lift in model performance when judged on a handful of different metrics. The EMR features did cause large shuffling in the ranking of predictions, so similar performance was reached with a noticeably different cohort. Possibly more important, the EMR features provided new and potentially more actionable reasons for a given patient's predictions.

Marginal effect estimates should likely be avoided when calculating and communicating the effects of individual features in this scenario; marginal effect estimates depend upon holding all other features constant. Given the highly overlapping and collinear nature of many of the features explored here, it is improper to even hypothetically hold all other features constant. Instead, reestimated univariate/single feature effects can communicate more useful information.

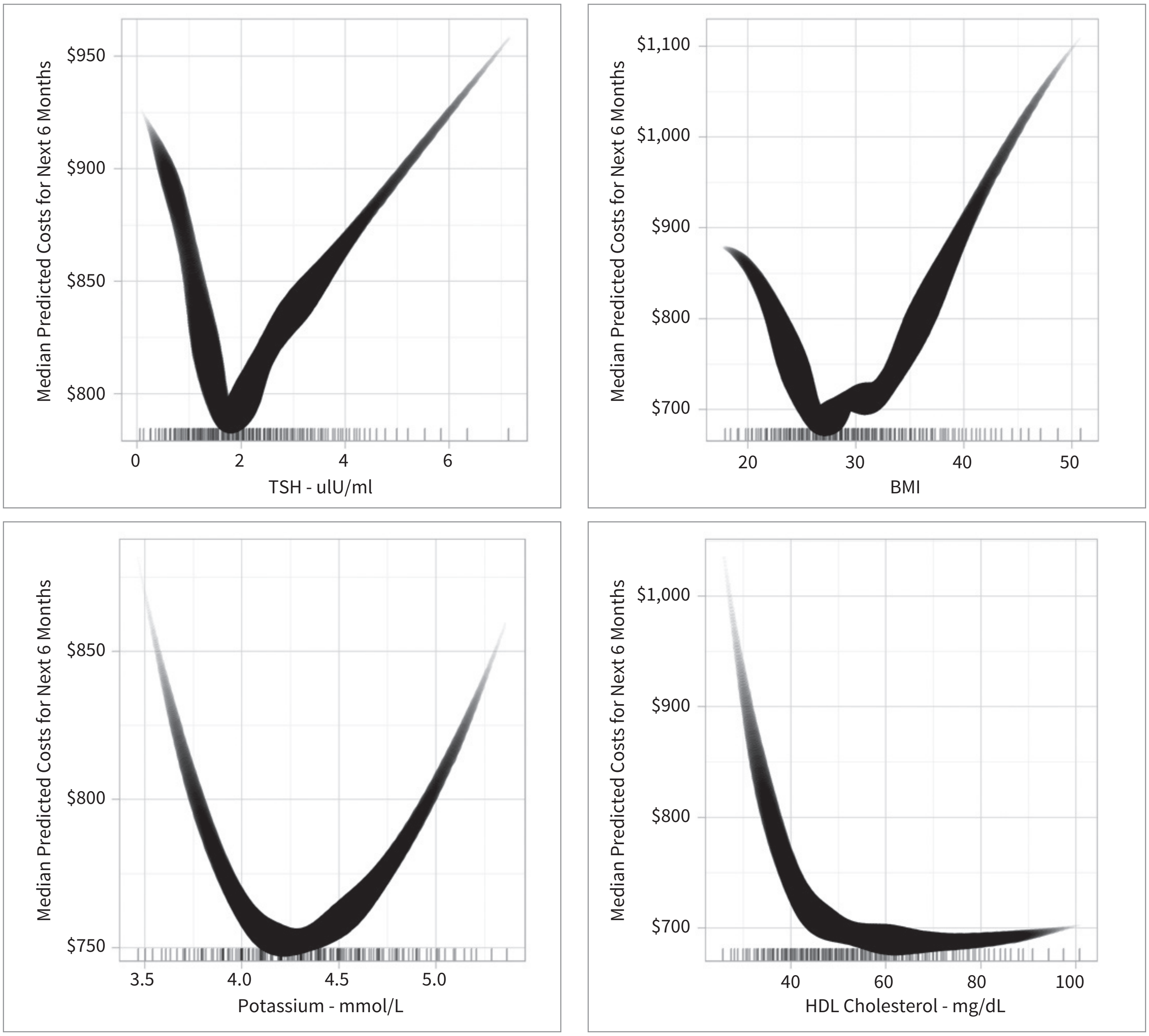
The reestimated relation between the median cost predictions and a few EMR features are shown in Figure 2. The rug plots and width of the lines emphasize the area of support that contains most of the example patients' results. The recurring horseshoe shape is very common in EMR effects and reflects a natural optimal equilibrium. These shapes also tend to align with general clinical guidance.

PRESENTING RESULTS

Care coordination can use these results for both their accuracy and their insights. The predictions themselves can help prioritize what patients are selected for care coordination. The insights can be presented to care coordinators in the form of individual patient profiles. Each patient profile presents many of the features for that patient and ranks them by their importance to the patient's overall prediction. Individual feature importance is derived from the reestimated effects presented in Figure 2 using a given patient's actual feature values. Labs and vitals that appear higher in the feature importance list can be especially valuable for care coordinators because they can represent more actionable information than just warnings of high historical utilization. Care coordinators could still go directly to an EMR for this informa-

Figure 2

Example Effects of EMR Information



tion, but this feature importance reporting puts the information in a useful context. Adding EMR information provided value, but more to inferential insights than predictive accuracy. However, the value of EMR information depends upon the process used to extract it and this only recounts one useful approach. ■

Sheamus Kee Parkes, FSA, MAAA, is an actuary at Milliman in Indianapolis. He can be reached at *shea.parkes@milliman.com*

Simple and Effective Reserve Practices: Approaches to Combine Your Reserve Estimates for Better Prediction

By Dale Cap, Chris Coulter and Kevin McCoy

The idea is simple—take the various predictions you are already making and weight them in a way that minimizes variance and increases accuracy.

Actuarial judgment is pervasive in our work. In many cases, judgment is a necessary element to our modeling and analysis. Over the past four decades behavioral research has shown that simple linear models can do much better than a human practitioner in many cases (Kahneman & Tversky, 2011; Wacek, 2007).

We present a couple of simple but effective reserving techniques that an actuary can add to his or her current reserving practices to produce significant reductions in reserve bias as well as reductions to reserve variance. Aggregating reserve estimates using only actuarial judgment can result in high variance and biased results, which can have consequences in many other areas of your company.

According to the Washington State Office of the Insurance Commissioner's data, the range of reserve error reported on financial statements for the largest insurance entities for the years 2008 through 2014, was –10 percent to 40 percent (Company Annual Statements, n.d.). More importantly, the standard deviation of these errors is 11 percent. This data supports the possibility of biases that actuaries generally believe to exist. Biases in reserve estimates include overcompensation (when you've reserved low one year, you overcompensate the next year by reserving way too high); or keeping too much weight on the prior estimates when new information is available; and more. It also indicates that the reserving techniques that are being employed are not very precise. With an 11 percent margin and an 11 percent swing, companies can easily see reserve estimates exceeding the final paid claims by up to 40 percent. This leaves capital in the prior year that could be used to benefit this year. This could impact the bottom line, distort the company's profitability over time, adversely affect ratings in the following year, trigger regulatory action, or impact pricing and forecasting models. Under-reserving can have similar effects. In addition to pricing and forecasting impacts, accruals may be set aside assuming a medical loss ratio (MLR) or other rebates are due, causing inappropriate payments on performance bonuses and bringing additional scrutiny to your department and deteriorating your credibility as the reserving actuary.

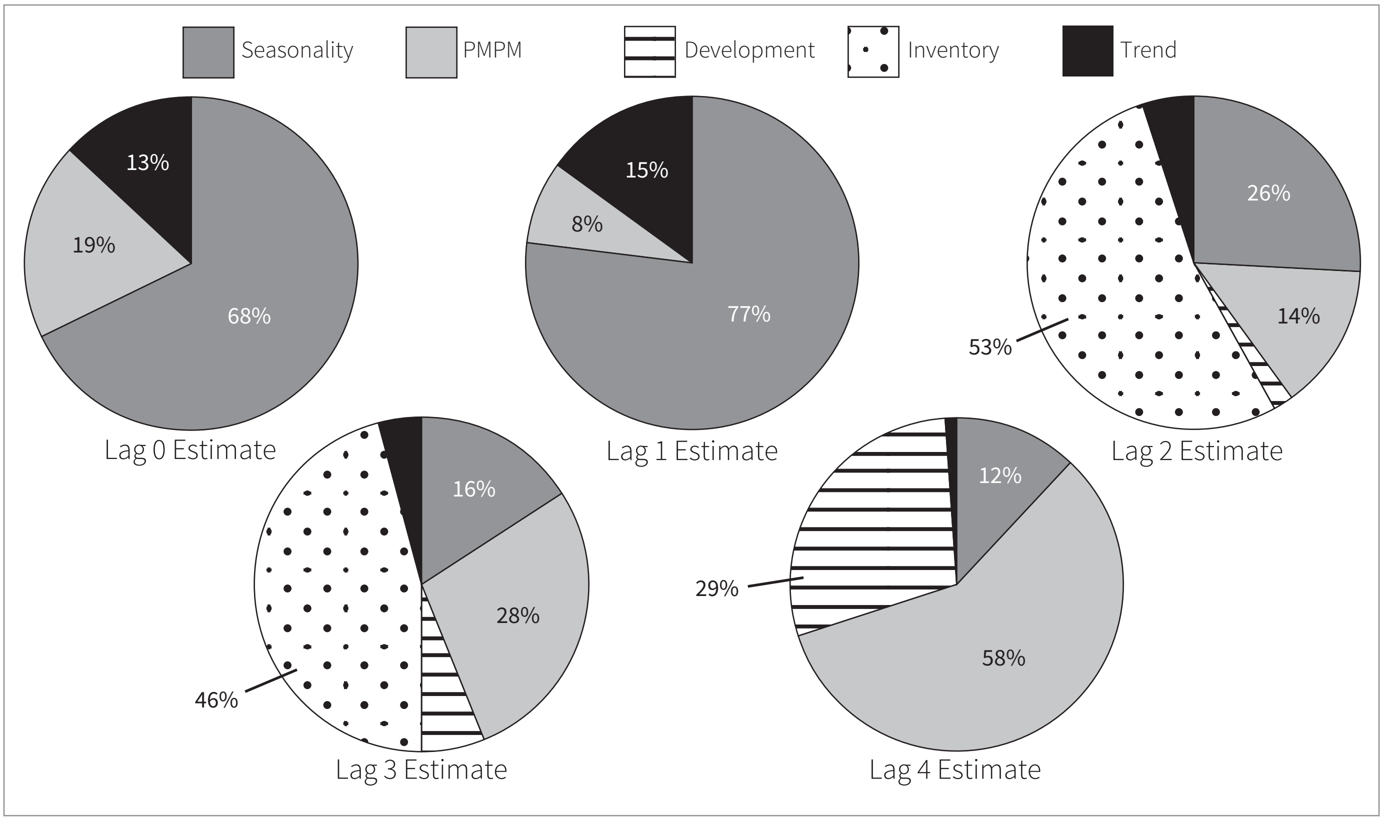
The following results are based on a simulation study with 8,000 simulations of claims run-out. The simulations took into account a seasonality component, a benefit change component, and a large claim component. Each of these components was developed with some randomness in each simulation. These simulations show a reduction of 5 percent variance to the reserve estimates. Unless estimators are completely correlated, these techniques should produce a reduction in variance and a more consistent estimate of the mean. With reduced variance and more accurate predictions, the margins needed could be reduced, resulting in a better estimate of each year's results.

The remainder of this article will outline the proposed techniques, followed by a high-level summary of the simulated data used to illustrate the results. Note: Although we illustrated the results by way of simulation, these techniques have been used in real practice and have shown a significant impact.

WEIGHTING TECHNIQUES

The idea is simple—take the various predictions you are already making and weight them in a way that minimizes variance and increases accuracy. This paper will discuss two weighting techniques you can use. However, there are many different ways to calculate the weights. Every reserving actuary is inherently doing this weighting in some fashion, whether it is via a mental algorithm or a more formalized approach. We advocate using a formalized approach that is testable and avoids potential human biases. In addition, the proposed formalized approach will tend to discredit reserving methods that perform poorly, focusing on those methods that are more reliable and consistent. If nothing else, this will give you a better baseline in which to apply judgment.

The following is an example illustrating the outcome from a weighting technique over multiple reserve methods by lag month.



In this example, we used the weighting technique to combine the seasonality, paid per member per Month (PMPM), development, inventory, and trend methods. As you can see each lag differs in the weights applied to each method. In Lag 0, the seasonality method had the highest weight, indicating that it was the "best" model for that lag. However, the seasonality method alone is not the best method. Rather, the weighting given in the above panel minimizes the variance of the estimate, so we would use that weighting for our predictions of Lag 0 claims.

We recommend ongoing monitoring and measurement of any approach used to ensure the intended outcomes and expectations are being met. One of the pitfalls of this more data-driven weighting approach is over-fitting. This is a common pitfall in any estimation or prediction procedure.

TECHNIQUE 1: INVERSE VARIANCE

Inverse variance weights each of the reserve methods based on the inverse proportion of error variance when comparing to actuals. Therefore, lower weights are applied to those methods that have historically produced a larger variation of errors.

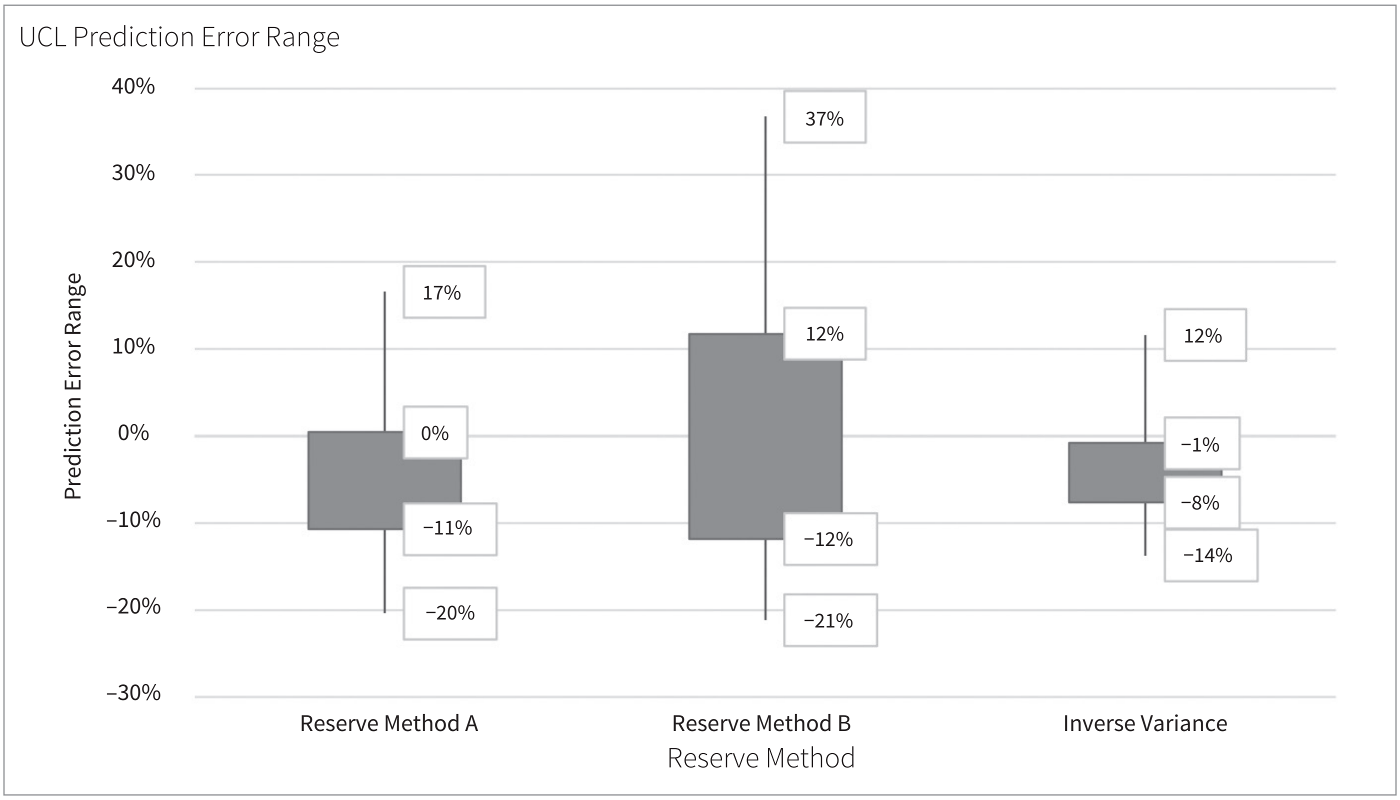
This approach is straightforward and simple to implement without having to add any additional features to one's existing reserve model. It also avoids any complex calculations, making it easy to explain to others. On the other hand, this type of approach ignores the correlations between the reserve methods being used and their distance from the target, which could be used to help lower the variance even further. This is why we offer two approaches.

Example

Suppose you have two methods for reserving: A and B. Each of these methods has a historical monthly reserve error associated with it (variance of 10 and 20, respectively). Based on the inverse variance technique, the proposed future weights when developing a projection could be 86 percent A and 14 percent B. This type of back-test has established that A is a better predictor; however, the mix of the two methods is still preferred. This technique provides a systematic approach to choosing a good mix and possibly better starting point prior to applying judgment in your reserve picks going forward.

Historical Experience	Method A	Method B	Actuals
Month 1	150.00	155.00	151.10
Month 2	160.00	145.00	155.20
Month 3	170.00	180.00	172.30
Variance of Monthly Errors	14.44	88.94	
Inverse Variance	0.07	0.01	
Proposed Future Weights	**0.86**	**0.14**	

After applying the inverse variance against our simulated claims database, using two of the more common reserving methods, we captured the unpaid claim liability (UCL) estimates for each incurred month. These estimates were then compared to the actual known liability, and their range of error is illustrated below. As seen below, the range of error using the inverse variance approach reduces the overall range of error when compared to each reserve method independently. However, you can also see that the technique doesn't improve accuracy significantly.



TECHNIQUE 2: LINEAR REGRESSION

The linear regression approach should produce more accurate weightings than the inverse variance approach, but it is far more computationally intensive. To ensure accuracy, the linear regression technique minimizes the sum of squared prediction errors for all points, penalizing larger errors disproportionately. On the other hand, the inverse variance focuses on reducing the dispersion of the estimates instead of the size of the error. In other words, the inverse variance method tends to enhance the precision of the estimate, but not necessarily the accuracy.

Example

Suppose you have two methods used for reserving: A and B. Each of these methods produced a historical estimate for the month. If we define A and B as X (a 2 x 3 matrix with A being column 1 and B column 2) and Y being the actuals, we could use the normal equation to solve for the proposed weights (assuming the matrix is invertible). Below is an example of the equation, where T is the transpose of the matrix and –1 is the inverse.

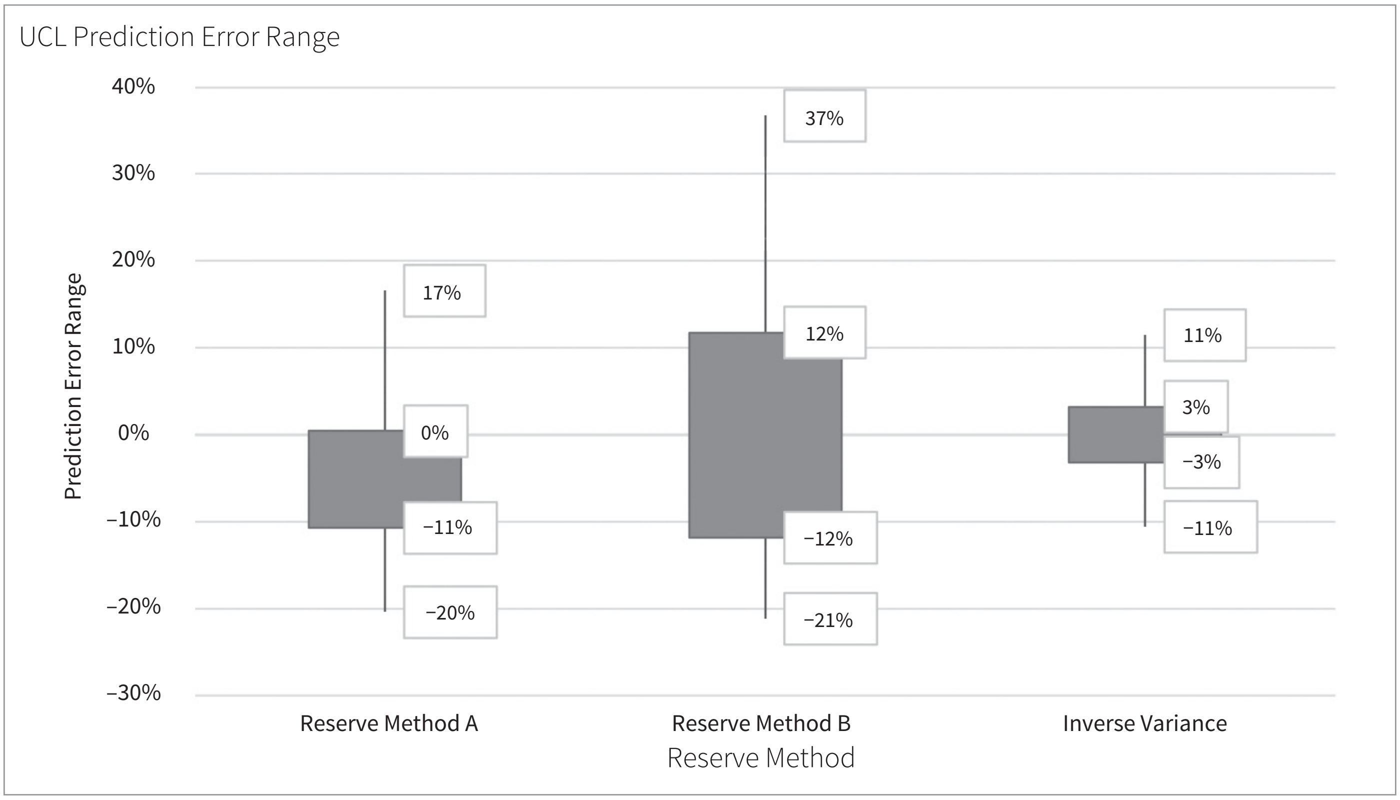
$$\text{Weights} = (X^T X)^{-1} X^T y$$

Applying this to the table below, the proposed future weights for these methods would be 71 percent A and 29 percent B (for this particular lag).

Historical Experience	Method A	Method B	Actuals
Month 1	150.00	155.00	151.10
Month 2	160.00	145.00	155.20
Month 3	170.00	180.00	172.30
Proposed Future Weights	**0.71**	**0.29**	

This type of back-test has established that A is a better predictor; however, the mix of the two methods is still preferable. This technique provides a systematic approach to choosing a good mix and possibly better starting point prior to applying judgment in your reserve picks going forward.

A similar illustration using linear regression against our simulated claims database can be found below. As discussed above, accuracy is what sets linear regression apart from the inverse variance approach. Unlike the previous results, the results here tend to center themselves on zero.



Although we provided an actual example where only two predictors are used, you can include more. Typically, actuaries may have many methods at their disposal like the development method, the paid PMPM method, loss ratio methods, trend-based methods, seasonality-based methods, etc. You can also integrate other variables into the analysis, such as the size of the current claims inventory. For whatever methods are ultimately chosen, we encourage you to pick methods that are diverse and not well-correlated with one another. We also encourage the methods to be consistent and stable over time. At the same time, you should be careful not to over-fit your data.

SUMMARY

In the examples outlined above, we presented two high-level techniques to weight existing reserve estimates. We showed how these techniques can improve your already defined reserving process with little extra work. In addition to the improvement to your estimates, there are two other benefits: the techniques will help the reserving actuary more precisely quantify where and when each reserving method works; and linear regression allows the actuary to integrate stochastic techniques in the calculation of reserve margins. However, there are limitations, and you should be aware of these and use judgment where necessary.

Predictive analytics is the practice of extracting information from existing data to determine patterns and predict future outcomes and trends (Predictive analytics, n.d.). If you don't use a weighting algorithm to combine your reserve estimates, you probably have a pretty good sense of which of your models performs the best for each lag month. But, the question is by how much. A weighting algorithm trained on real data can give you more precision around which models work better and when.

"Predictive analytics" is the new catch phrase, but not long ago stochastic analysis was a hot topic. Reserving is certainly a place where more stochastic models can prove beneficial. A Society of Actuaries sponsored report gives a definition of what margin is for incurred but not reported (IBNR). In math, it is written as:

$$\text{Probability(Estimate + Margin} > 95\%) > 85\%$$

The report also gives the reader a couple of ideas on how to obtain this estimate (Chadick, Campbell & Knox-Seith, 2009). In this report, the authors also point you to another Society of Actuaries published report, *Statistical Methods for Health Actuaries IBNR Estimates: An Introduction*, which outlines some more sophisticated ways to statistically approximate your IBNR (Gamage, Linfield, Ostaszewski & Siegel, 2007). Using Technique 2 is a great first step in integrating the stochastics into your already defined reserving system.

The idea of combining two or more estimates for better prediction or lower variance is used in many other contexts; it's called meta-analysis in statistics and ensemble methods in data science, while in finance the capital asset pricing model (CAPM) uses an optimal weighting structure. In any case, they work and can help to reduce the biases that exist in your reserving process.

DATA AND SIMULATIONS

Although these techniques have been shown to be successful in practice, the results included in this paper were developed using data from our simulated claim database to avoid the use of actual data in this paper. The ultimate incurred claims were developed by lag month and include adjustments for changes in claim processing patterns, number of weekly paid claims in a month, benefit design, workday factors, random large claim shocks, seasonality, leveraging, and other factors (which include random noise within each component and overall).

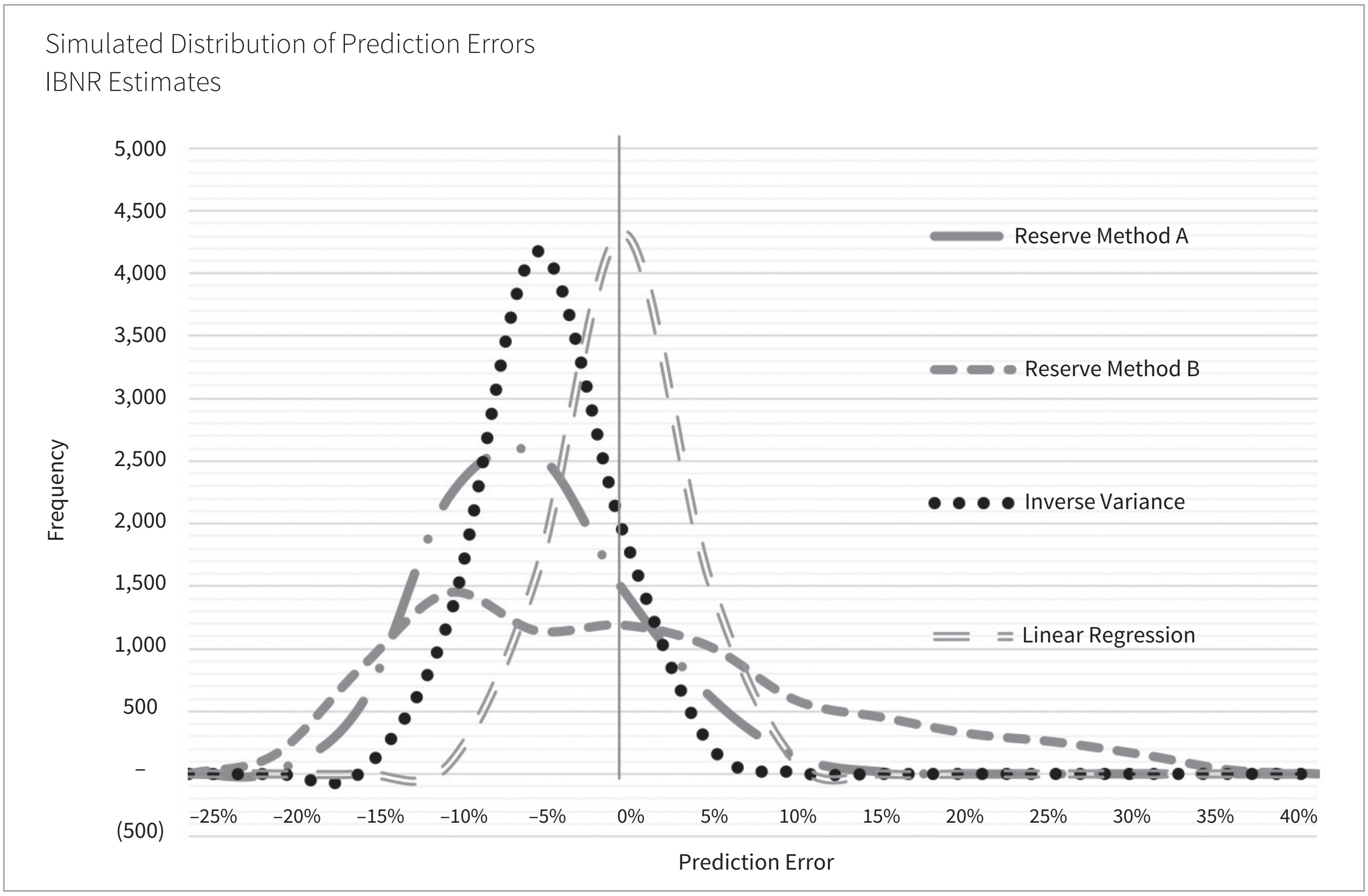
Consistent with actual experience, our simulated examples have shown improved performance when compared to using a single method for reserving. Although we are not able to simulate judgment, we have seen actual improvement when comparing to our final picks (adjusting for margin and implicit conservatism), but we will leave it to the readers to test their own historical performance and whether these techniques add value (or just a better baseline from which to build their estimates).

In the end, we believe if employed correctly—using various reliable and stable methods—these techniques (particularly regression) can help reduce both the bias and variance in the estimates.

Below are the results obtained from applying these techniques to our claims database. Roughly 8,000 simulations were generated estimating the ultimate claim liability for a given month.

Summary Statistics				
Statistics	Reserve Method A	Reserve Method B	Inverse Variance	Linear Regression
Mean Error	−5.2%	0.0%	−4.2%	0.0%
Std Error	5.6%	11.8%	3.4%	3.2%
Kurtosis	−1.9%	−14.8%	24.1%	−0.8%
VAR95%	13.5%	23.4%	9.36%	6.3%
Skew	43.2%	68.3%	30.8%	2.4%
Worst Error	**20.4**%	**36.7**%	13.7%	11.5%

VAR95% represents the point at which 95 percent of the errors (in absolute terms) fall below.



REFERENCES

Chadick, C., W. Campbell and F. Knox-Seith. (2009). *Comparison of Incurred but Not Reported (IBNR) Methods*. Society of Actuaries and Lewis and Ellis.

Company Annual Statements. (n.d.). Retrieved from Washington State Office of the Insurance Commissioner: *http://www.insurance.wa.gov/for-insurers/look-up-companies-agents/ins-company-annual-statements/.*

Gamage, J., J. Linfield, K. Ostaszewski and S. Siegel. (2007, Oct. 31). *Statistical Methods for Health Actuaries—IBNR Estimates: An Introduction.* Retrieved from Society of Actuaries: *https://www.soa.org/research/research-projects/health/research-stats-hlth-act.aspx.*

Kahneman, D., and A. Tversky. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.

Predictive analytics. (n.d.). Retrieved from Wikipedia: *https://en.wikipedia.org/wiki/Predictive_analytics.*

Wacek, M. (2007). *A Test of Clinical Judgement vs. Statistical Prediction in Loss Reserving for Commercial Auto Liability.* Casualty Actuarial Society.

Dale Cap, ASA, MAAA, is a director at Aetna in Seattle and can be reached at *capd@aetna.com.*

Chris Coulter, FSA, MAAA is an actuary at Cambia Health Solutions in Seattle and leader of the Behavioral Finance Subgroup of the SOA's Health Section. He can be reached at *actuarychris@gmail.com.*

Kevin McCoy is a senior actuarial analyst at Cambia Health Solutions in Seattle. He can be reached at *Kevin.Mccoy@regence.com.*

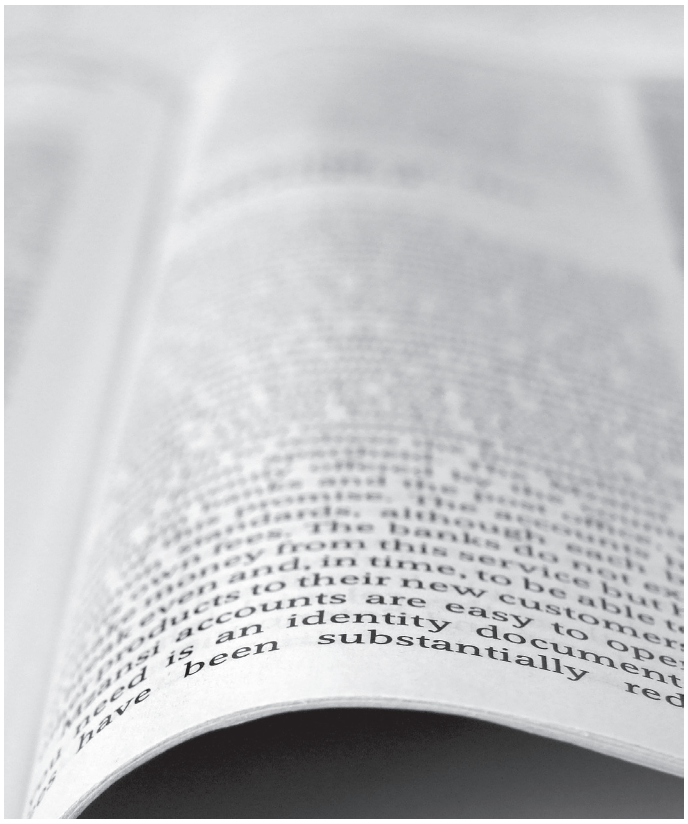
Articles in the *North American Actuarial Journal* of Interest to Health Actuaries

By Ian Duncan

After a good run of health-related articles, Volume 19, No. 2 did not have any articles of direct interest to health actuaries, although those actuaries interested in the relationship between disease and longevity may be interested in "Causes-of-Death Mortality: What Do We Know on Their Dependence?" by Séverine Arnold (-Gaille) and Michael Sherris. (We reproduce the abstract of this article below.) Health actuarial topics are back in force in Vol. 19, No. 3 with several interesting articles. Colin M. Ramsay and Victor I. Oguledo address an increasingly important topic for health actuaries—absenteeism and presenteeism—in "Optimal Disability Insurance with Moral Hazards: Absenteeism, Presenteeism, and Shirking." Sam Gutterman has an article on "Mortality of Smoking by Gender," which I am sure all health actuaries will want to read. The relationship between health and longevity continues to be an important source of articles with "Mortality, Health and Marriage: A Study Based on Taiwan's Population Data" by Hsin Chung Wang and Jack C. Yue.

ABSTRACT: CAUSES-OF-DEATH MORTALITY: WHAT DO WE KNOW ON THEIR DEPENDENCE?

Over the last century, the assumption usually made was that causes of death are independent, although it is well-known that dependencies exist. Recent developments in econometrics allow, through Vector Error Correction Models (VECMs), to model multivariate dynamic systems including time dependency between economic variables. Common trends that exist between the variables may then be highlighted, the relation between these variables being represented by a long-run equilibrium relationship. In this work, VECMs are developed for causes-of-death mortality. We analyze the five main causes of death across 10 major countries representing a diversity of developed economies. The World Health Organization website provides cause-of-death information for about the last 60 years. Our analysis reveals that long-run equilibrium relationships exist between the five main causes of death, improving our understanding of the nature of dependence between these competing risks over recent years. It also highlights that countries usually had different past experience in regard to cause-of-death mortality trends, and, thus, applying results from one country to another may be misleading.



OPTIMAL DISABILITY INSURANCE WITH MORAL HAZARDS: ABSENTEEISM, PRESENTEEISM, AND SHIRKING

Presenteeism occurs when employees are present at the workplace but cannot perform at their best due to ill health or other reasons, while absenteeism occurs when employees are absent from the workplace. While absenteeism is important, researchers now say presenteeism can be more costly to businesses and may be responsible for as much as three times the health-related lost productivity as compared to absenteeism, and may cost the U.S. economy as much as US$150 billion per year. Given the cost of absenteeism and presenteeism, one of the objectives of this paper is to provide actuaries with the techniques and insights needed to design disability insurance policies that take into account the dynamics of absenteeism and presenteeism. To this end we develop a simple multi-state sickness-disability model of the evolution of an employee's health over time. We assume employees receive sick-pay, the size of which depends on their health state and there is a government-sponsored unemployment insurance program. In our model it is possible for employees in good health to avoid work by staying home, which is called *shirking*. To reduce shirking, the employer decides to check the health status of a certain percentage of employees who call in sick. Given the sick-pay structure, the probability of a health check, and the existence of unemployment insurance, employees develop rational strategies about whether to engage in shirking, absenteeism or presenteeism. These strategies are captured in a set of Volterra integral equations. We use these Volterra integral equations to show how the employer can design a disability insurance plan that can incentivize employees to eliminate shirking and to act in a manner that will maximize the employer's expected profits.

MORTALITY OF SMOKING BY GENDER

Exposure to cigarette smoke has had and will continue to have a huge effect on mortality. Significant differences in smoking prevalence rates by gender have contributed to varying levels and rates of improvement in mortality over the last several decades and are expected to continue to influence mortality improvement differently over the next several decades.

The combined effect of greater historical smoking prevalence rates by males and their corresponding earlier and larger reduction has in part been responsible for the recent improvement in mortality rates for males compared to that for females in the United States. Similar patterns are evident in almost all economically developed countries, although their timing and levels differ. The patterns in less-developed countries will likely follow similar patterns as concerns emerge about the effect of smoking on the mortality of their citizens.

The objective of this paper is to compare smoking prevalence and cessation by gender and the effect on smoking-attributable and, in turn, all-cause mortality. A summary of mortality attribution approaches used to enhance the evaluation of the effect of smoking and projections of mortality rates by gender is also provided. ■

Ian Duncan, FSA, FCIA, FIA, MAAA, is adjunct professor of actuarial statistics at the University of California, Santa Barbara. He can be reached at *duncan@pstat.ucsb.edu.*