SOCIETY OF ACTUARIES

Article from:

Health Section News

August 2003 – Issue No. 46

SOCIETY OF ACTUARIES

Issue No.46 | August 2003

Health Section News

"For Professional Recognition of the Health Actuary"

Applying Diagnosis-Based Predictive Models to Group Underwriting

by Randall J. Ellis, Marilyn Schlein Kramer, Joseph F. Romano and Rong Yi

Introduction

The development of group premium health rates is a joint effort of pricing actuaries and underwriters. Typically actuaries set the broad rating approach and factors while underwriters examine the realities of a specific group against the theoretical relationships and guidelines inherent in the approach. Traditional actuarial tools and developments are used to arrive at such approaches and factors, but all are ultimately aimed at one point ... the understanding of risk associated with the specific group.

These traditional actuarial approaches and underwriting guidelines use demographic information such as age/sex, type of occupation, financial stability, insurance carrier turnover, employee turnover and prior cost experience to analyze risk. However, there is significant data supplied in medical claims that can improve the match of premium to expected medical expense, improve group retention and ultimately improve long-term financial results, assuming one can properly examine the data to better predict the implications of such experience use.

Predictive models that use medical and pharmacy claims information to accurately measure expected health care

consumption to support efficient allocation of resource have been the subject of increased interest. Most studies have focused on the statistical predictive power of the models, (i.e. R-squared values on the standard measure of a model's predictive power).[1]

As indicated in the study noted, diagnosis-based models outperform age/sex approaches. Coupled with the improvement in data accessibility and

(continued on page 4)

1 See the recent SOA research publication "A Comparative Analysis of Claims-based Methods for Health Risk assessment for Commercial Populations, May, 2002.

quality over time, the cost to process such data as well as the cost of the models to use such information are small relative to the possible savings from improved margins.

However enticing such results might be, the prospective user of any predictive model approach for rating will want to review the answers to several key questions:

- How do diagnosis-based models compare against commonly used group underwriting models?
- What is the increased benefit (return on investment) from the added predictive power of the diagnosis based models?
- How can these models be used in the real world of a health plan which renews 100s to 1,000s of accounts each month?

In this study, we

- Compare a predictive model approach with traditional experience rating approaches. In performing this comparison we used the DxCG predictive models to examine the impact at various "group size" levels.
- Describe a new methodology for assessing model performance in group underwriting by using economic modeling principles, including simulation studies. The concept of "actuarially balanced rating" or "actuarially fair" rates as used in this simulation is described below.
- Suggest areas of further study and collaboration by health services researchers and actuaries.

Actuarially Balanced Rating

Most insurers break the universe of groups into broad categories, such as small group, mid market and large group. While the actual points of size delineation vary from insurer to insurer and geography to geography, the fundamental reason for such categorization is regulatory constraints which typically affect the amount or degree to which actual group data can be used to rate that specific group. This limitation can be extreme at the smallest of group sizes, but generally eases significantly as group size increases, linking expectedly with standard actuarial understanding of credibility metrics and group size pooling. Regardless of any limitation on use of information, however, understanding of a group's cost expectations is vital to efficient and optimal planning and profit realization at all group size levels.

Just as important, is a balancing of the various aspects that come together to develop a reasonable rating approach. While different actuaries may use slightly different terms, the essences of actuarial balance as used in this paper can be summarized as an appropriate blending of competing factors, any one of which, in the extreme, can lead to undesirable financial results. For purposes of understanding the simulation's use of balanced or "fair" rates, we can assume that the rates developed must be adequate to cover expected costs plus expenses and other profit loads, be competitive, be reasonably simplistic for both internal and external understanding, have a level of flexibility to respond to emergent issues and be compatible with necessary provider and regulatory constraints. For our purposes, rating structures which appropriately blend or balance these factors will result in rates that can be considered "fair" for all parties involved (i.e., the insurers and the group).

For the models, data and methods, we use the Diagnosis Cost Group (DCG) risk adjustment model.[2] The prospective DCG model uses a year of medical claims data (and demographics) to predict next year's costs at the individual and group levels. In predicting next year's costs, the DCG model identifies chronic conditions (that predictably and systematically result in higher costs) and quantifies their impact. Non-chronic conditions, such as broken legs, pneumonia, etc., are not used to predict costs because there is no reason to believe that having pneumonia in year 1 is associated with higher costs in year 2.

Specifically, we use the DCG/HCC (Diagnostic Cost Group/Hierarchical Condition Category) model which uses diagnoses from all sites of service—inpatient and outpatient. The DCG/HCC model is the basis by which Medicare will pay Medicare+Choice plans beginning in 2004. We refine the DCG/HCC model by also using prior costs and refer to it herein as the " DCG Underwriting Model."

We compare the DCG Underwriting Model to a traditional underwriting model. Since underwriting models tend to be proprietary, we used an age-sex, prior cost (experience rating) model as proxy for the "Traditional Model."

Thus, the sole difference between the DCG Underwriting Model and the Traditional Model is the use of diagnoses in the DCG model.

Since information on actual market prices (premiums) offered and accepted by employers are

2 Cite paper with DCG model description. Also refer to DxCG Web site.

unavailable for analysis, we use quantitative analyses based on simulations of employer groups drawn from the MedStat MarketScan commercial dataset from years 1997 through 1999. The dataset has demographic, diagnostic and pharmacy and cost information on over 2.38 million members who were eligible for health care insurance for at least 1 month in both 1997/98 (Year 1) and 1998/99 (Year 2). The dataset is drawn from employer-sponsored health plans from across the country. The data include fee for service, PPO and HMO plans and various benefit levels.

Unlike consumer products where market researchers can adjust prices and directly observe the consequences in terms of consumer demand and profit, the health care market offers no such "laboratory" for testing prices. As a result, we develop a "bidding system" to assess the impact of the models in terms of number of group accounts secured and resulting margins or profits in winning accounts.

We use a fixed sample size of 500,000 randomly selected from the MarketScan dataset. To simulate employer groups, random draws (with replacement) were used.

For each group, expected costs were calculated using the DCG Underwriting Model and the Traditional Model. Load factors (margin) were added to both to obtain "bid prices." To simplify the analysis, we did not consider the impact of state underwriting regulations. The lower price "wins" in each bidding process and market share is defined as the percentage of bidding processes won. Profitability is calculated as the difference of the "winning" price and actual Year 2 expenses for the group.

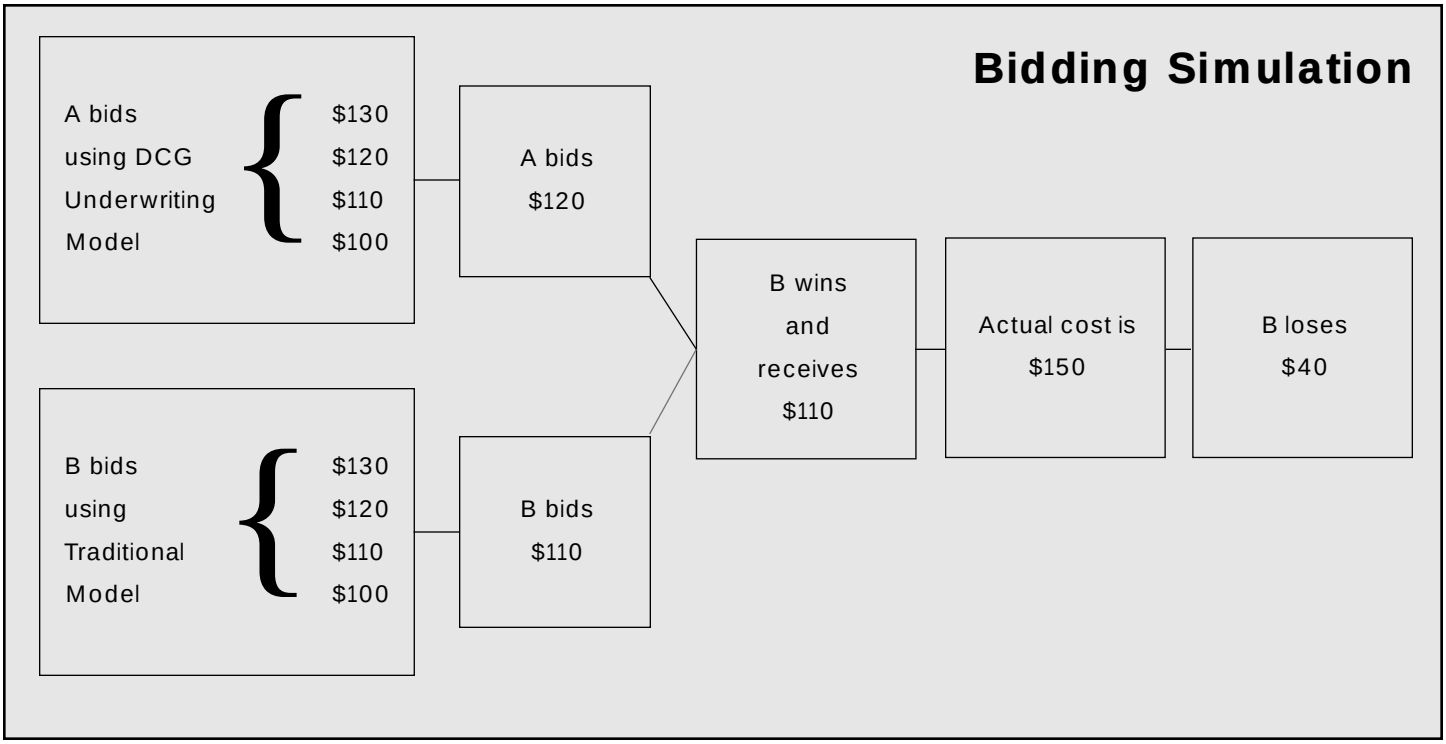
The following graph depicts the bidding process. Using the DCG Underwriting Model and diagnoses from Year 1, Insurer A bids $120 for year 2 while Insurer B using the Traditional Model bids $110. Insurer B "wins" having bid $10 less than Insurer A. Insurer B wins and receives $110. However, the actual (Year 2) per month cost of that group is $150 so Insurer B loses $40 per person.

The same bid calculations and bid results were simulated on group-sizes of 5, 25, 50, 100 and 500 lives with varying load factors. For each run, bids and results (market share, or number of accounts won, and profitability) for 1,000s of groups were calculated.

Results

Table 1 presents that individual R2 for each of the models. Note that these are validated R2 statistics, meaning that the models were calibrated on one set of data and test (validated) on a second set of data.

(continued on page 6)



Top-coding refers to capping claims levels at various thresholds ($25,000, $50,000 and $100,000 per person in a year). For example, a person with $125,000 in claims would be top-coded at $100,000 if a $100,000 threshold were imposed.

How and where does the DCG Underwriting Model outperform the Traditional Model? Table 2 compares the confidence intervals (CI) and positive predictive values (PPV) at each small group size between the two methods. We see that the DCG Underwriting Model has smaller confidence intervals at all group sizes, and this advantage is more pronounced as group size becomes smaller. In other words, the DCG Underwriting Model is more accurate than the Traditional Model in getting the group mean expenditure right. When looking at the high-cost or low-cost end of the population, we can see that the DCG Underwriting Model identifies more people when they are truly high or low cost at all group sizes.

Thus, the DCG Underwriting Model outperforms the Traditional Model in all key aspects. The question remains however, how much is that added predictive power "worth" when the models are used in underwriting?

Table 3 on page 7 answers this by presenting the bidding results and compares them accordingly to each small group size. Here we compute the number of accounts won by each method and their profits by assuming that the Traditional Model is fixed at 10 percent load factor while the DCG Underwriting Model varies between 0 to 10 percent. For example, when both assume 10 percent load factor, the DCG Underwriting Model generates $146 profit per member per year, while the Traditional Model generates $7 loss per member per year. As the DCG Underwriting Model lowers its load factor, it wins more accounts, although at a lower profit.

Table 1: Validated Individual Level R2 Statistics

	Traditional Model (Age/Sex and Prior Cost)	DCG Underwriting Model (DCG and Prior Cost)
Not Top-coded	11.4%	20.9%
Top-coded at $100k	13.7%	26.9%
Top-coded at $50k	13.0%	28.5%
Top-coded at $25k	11.7%	31.1%

Table 2: Confidence Interval and Positive Predictive Value

Group Size	Model	CI-to-Mean Expenditure (%)	Top 20% PPV	Bottom 20% PPV
100	DCG	4.97%	42.7%	42.5%
	Traditional	5.23%	39.0%	38.0%
50	DCG	9.21%	42.4%	43.6%
	Traditional	9.80%	39.2%	39.0%
25	DCG	16.41%	44.5%	41.7%
	Traditional	17.51%	39.6%	39.9%

Table 3: Comparison of Results with Traditional and DCG Underwriting Models

Group Size	Load Factor	% of Accounts "Won" Using DCG	DCG Underwriting	Traditional
100	10%	40.5%	$146	($7)
	9%	44.6%	$132	($22)
	8%	48.1%	$116	($35)
50	10%	47.3%	$92	($58)
	9%	49.9%	$84	($760)
	8%	52.4%	$73	($92)
25	10%	50.3%	$84	($138)
	9%	52.3%	$70	($152)
	8%	54.4%	$55	($164)

* Note: Assume constant load factor of 10% for the Traditional Model.

At all group sizes studied and for each load factor, the DCG method outperforms the traditional method in terms of profitability. DCG wins more accounts at the smaller group level.

Impact of Results for Underwriters

These results show that adding diagnosis information significantly improves predictive power of traditional methods with increased accuracy and specificity thereby supporting the goals of an actuarially balanced rating system. These findings support our understanding of how "one-time" and chronic conditions impact historic and future costs. Diagnosis models allow us to appropriately adjust for the impact of one-time conditions (broken legs, pneumonia, etc.) from future cost predictions while prior cost models implicitly assume that the high costs in Year 1 are "rolled forward" to Year 2. Conversely, someone diagnosed with metastatic cancer based on a diagnosis late in Year 1, will have very high costs predicted for Year 2, even if claims experience in Year 1 were relatively low. Such future costs would likely be understated in the traditional model approaches.

Moreover, as can be seen by the modeling, the improvement in predictive power translates into superior margins for health plans incorporating diagnoses into their underwriting models. Assuming the full costs of licensing and implementing diagnosis predictive models is charged to rating, net margins will still be higher than using traditional methods only when such costs are spread over the entire rating pool. Since such predictive models are also useful for medical management, a broader and lower "rating" allocation is reasonable, thus increasing the positive margin improvement when using a predictive model approach.

Some Final Thoughts

The paper has presented a preliminary examination about the natural extension of predictive software for use in a group underwriting environment. Although regulations may effectively limit a complete application of the methods described for certain size groups, the overall results will still likely be net accretive to the bottom line.

Like many approaches in rating, the methods indicated here may best be integrated over a period of time (for example, blended by "credibility like" factors to stabilize any year to year movement in a particular group's renewal levels).

And although the approach described here was focused upon renewal underwriting, use for new group rating points or debit approaches are possible and desirable. For example, insurers who use rating approaches that employ such point/debit systems generally determine such points based upon new enrollment questionnaires. The predictive model approach can validate, refine and expand the scientific basis of their point assignment to more accurately reflect the impact of one time versus recurring costs. Such a refinement would also support a new group/renewal linkage and bridge the transition of new group and renewal rating.

Randall J. Ellis, Ph.D., is a professor of Economics at Boston University in Boston, MA. He can be reached at 617-353-2741.

Marilyn Schlein Kramer, works for DxCG, Inc. in Boston, MA. She can be reached at 617-303-3790.

Joseph N. Romano, ASA, MAAA, is a consulting actuary in Thousand Oaks, CA. He can be reached at seeandbasett@msn.com.

Rong Yi, Ph.D, also contributed to this article.