SOCIETY OF ACTUARIES

Article from:

Health Section News

August 2005 – Issue 50

Optimal Small Group Renewal Methods

by Ross Winkelman

Ask 10 actuaries to describe the best methods for rating small groups at renewal, and you are likely to get 10 different answers. Milliman surveyed 20 small group carriers on their renewal methods and found that they did things very differently. In particular, some plans used risk adjusters while many others did not. This variation indicated a critical need in the marketplace to identify the optimal small group renewal methods and to quantify the value of implementing those methods. Milliman is in the midst of a research study to identify optimal small group methods (The study will be completed as of the date this article is published.) To identify the characteristics of the optimal methods and quantify their value, we are answering a number of important questions:

1. What is the most accurate way to rate small groups at renewal using traditional information and methods?
2. How much do risk adjusters improve precision under real-world conditions?
3. Should risk adjusters be combined with traditional loss ratio approaches? How should risk adjusters be calibrated?
4. Does credibility really increase with group size?
5. What other factors affect credibility?
6. What is the bottom line impact of improving renewal methods?
7. How does what your competitors are doing affect what you should do?

To date, we have reached some surprising conclusions. We have concluded that risk adjusters significantly improve precision, but only marginally do so under real world conditions. This is especially apparent as group size increases and state limits on allowable rate variation due to health status shrink. Risk adjusters should be calibrated for the specific block of business being rated and should be combined with traditional methods to optimize their precision. Traditional information and methods can be optimized to a point where they perform well when compared to methods that use risk adjusters. Finally, the value of improving renewal methods was lower than we originally expected, especially considering that renewal methods compete against new business methods, not other renewal methods.

There are many potential reasons for using risk adjusters in small group renewal rating, including the following:

1. Risk adjusters were developed to predict morbidity, and numerous studies have proven their effectiveness.
2. Risk adjusters can be calibrated to the specifics of a block of business (i.e., relative costs for different conditions, provider networks, benefit design, etc.).
3. Risk adjusters can use prescription drug data by itself and still produce good morbidity estimates. This is a very interesting characteristic as claims completion is not available when a renewal is prepared and prescription drug claims run out very quickly.

Risk adjusters are not free, and licensing costs can be significant. In addition, implementing any new rating methods, including a risk-adjusted small group rating methodology, requires resources. Carriers need to weigh the costs and benefits of any new rating methodology, with the specifics of their block of business in mind.

Study Methods

Last fall, Milliman partnered with two small group carriers and launched a study to identify the optimal methods to renew small groups. We focused on risk adjusters in addition to studying the best way to use all of the information available at renewal, such as historic claims. Our study was performed with real-world conditions in mind, including state rating limits, which can vary substantially. For example, California allows a rating adjustment of only ± 10 percent due to health status, while Idaho allows ± 50 percent. Our error values were calculated relative to allowed rating action. For example, assume two different renewal methods estimate the health status of a group at 1.50 and 1.55. Also assume that the allowable rate variation in the state is only ± 35 percent and actual claims turn out to be 40 percent above average due to health status. In this example, we would identify both methods as producing the optimal answer—they both would assign a health status for this group as high as possible. In a statistical measurement for this case, we would set the mean absolute prediction error (MAPE) equal to zero and we would also set the sum of squared errors to zero (this translates to an R-squared of 1.0).

The other important real-world conditions faced by health plans include turnover and the fact that they are competing against other carriers' new business methods, not their renewal methods. Members active

Ross Winkelman, FSA, MAAA, is principal and consulting actuary at the Denver health office at Milliman. He can be reached at 303.672.9059 or *ross.winkelman@milliman.com.*

during the historic period are not necessarily the ones who will be active during the rating period. Further, even the members that are active at the time the renewal quote is developed may not be around during the entire rating period. This condition dampens the value of historic information for developing projections, because information for only a portion of members is available. We analyzed one set of carrier data and found that approximately 15 percent of the members active during the rating period did not have complete data for the experience period immediately prior to the renewal (eight or nine months of data for first renewal).

Previous studies have quantified the impact of using risk adjusters without considering state allowable rate variation limits. In addition, previous studies have simulated market conditions by pitting renewal methods directly against each other. Under actual market conditions, renewal methods compete against other carriers' new business methods. This is an important distinction. Typical new business methods are significantly less predictive than renewal methodologies, and the distribution of new business predictions is very different than the distribution of renewal business predictions.

We used claims, membership and premium data from a small group carrier, along with their manual rating formula in our study. In addition, we used a large set of individualized data from a variety of carriers. The individualized data did not include manual rates or actual small groups. The carrier information was used to study results under real - world conditions, while the individualized data, with its larger size, was used to measure the impact of relatively small changes in methodology. We were particularly careful not to calibrate methods on the same set of data against which we were testing. Our calibration steps were designed to mirror the approach a carrier would use to calibrate their methods. (We used different time periods for calibration and testing, instead of splitting data for the same time period into calibration and testing pieces. Splitting the data would require more data than most carriers have access to.)

Study Results

We compared the PMPM MAPE expressed as a percentage of the claims PMPM for traditional methods and for traditional methods supplemented with risk adjusters. MAPE is calculated as the absolute value of the difference between PMPM predicted values and PMPM actual values. Smaller MAPE PMPM percentage values indicate better precision. MAPE values decrease with smaller allowable rate variation, because there is less variation to predict. In addition, MAPE values decrease as group size increases because costs approach the mean as group size increases.

The following table shows MAPE PMPM percentage results for the first renewal using the individualized data. We also studied the marginal improvement in prediction at second renewal and the results were similar.

MAPE (as %)

Methodology	1 Member Uncapped	10 Employees Uncapped	10 Employees ± 35%
Manual Rate	101.02%	36.81%	25.53%
Traditional Methodology	90.75%	33.89%	22.51%
Risk Adjuster Methodology	82.67%	31.95%	20.86%

The methodology used to calculate the error values above is a traditional loss ratio approach, applied in an optimal way (meaning with credibility weights that minimize the sum of squared error and other values calculated appropriately). The risk adjuster methodology uses both risk adjusters and traditional methods.

Implementing any new rating methods, including a risk adjusted small group rating methodology, requires resources.

As shown in the previous table, the MAPE is smallest for the risk adjuster method. Also, the MAPE results are very similar for all three methods (method one being a manual rate without any experience adjustments), especially with rating limits and for larger groups. The distribution of predictions compared to actual results presented later in this article helps understand this further.

R-squared measures the percentage of the variation from the mean that is explained by the rating methodology. An R-squared of one indicates that the method explains all of the variation from the mean, while an R-squared of zero indicates that the method does not explain any of the variation from the mean. Therefore, greater R-squared values indicate better precision. The following table shows preliminary R-squared results for the first renewal (second renewal results were similar).

R-Squared

Methodology	1 Member Uncapped	10 Employees Uncapped	10 Employees ± 35%
Manual Rate	0.0571	0.0438	0.1617
Traditional Methodology	0.1645	0.1638	0.2779
Risk Adjuster Methodology	0.2408	0.2527	0.3081

As shown in the R-squared table, the risk adjuster methodology outperforms the traditional methodology for one-employee groups without rating limits, and for larger groups. As the group size increases, there is less variation from the mean to explain. However, the risk adjuster methodology still does a better job of explaining that variation than traditional methods.

We also tested using traditional information by place of service (inpatient, outpatient, prescription drug). Our results are very promising and some error measures actually show this methodology outperforming methods that use risk adjusters. This approach results in a low credibility weight for historic inpatient costs, and relatively high weights on historic outpatient and prescription drug costs.

Another way to compare methods is to look at how well they assign groups into their cost categories—below the allowable rating variation, within the rating variation and above the rating variation. Assigning members into broad cost categories is a strength of risk adjusters and we expected them to perform well using this measure.

The following grid shows the nine possible combinations of actual and predicted costs relative to the allowable rating variation with the x-axis being predicted costs, and the y-axis being actual costs (the example illustrated is relative to ±35 percent allowable rating variation). All errors are expressed as a percentage of base rates:

ACTUAL \ PREDICTED			
	Actual > 1.35. Predicted < 0.65. Error = 70%	Actual > 1.35. Predicted within 0.65 to 1.35 Error ≈ 35% on Avg	**Actual > 1.35 Predicted > 1.35 Error = 0%**
	Actual within 0.65 to 1.35 Predicted < 0.65 Error ≈ 35% on Avg	Actual within 0.65 to 1.35 Predicted within 0.65 to 1.35 Error = Actual – Predicted	Actual within 0.65 to 1.35 Predicted within 0.65 to 1.35 Error ≈ 35% on Avg
	Actual < 1.35 Predicted < 0.65 Error = 0%	Actual < R.L. Predicted < R.L. Error ≈ 35% on Avg	Actual < 0.65. Predicted > 1.35 Error = 70% Predicted

For example, the top middle cell in the grid above describes a situation where the carrier estimates that costs will be within the allowable rate variation (between 65 percent and 135 percent of manual rates), but costs actually turn out to be greater than the allowable rate variation (more than 35 percent above manual rates).

In the same grid, we have included the MAPE (or a rough estimate of the average MAPE) specific to each cell. For example, in the upper left cell, actual costs are greater than the allowable rate limit, while the prediction would be made below the allowable rate limit. Therefore, the Mean Absolute error in this instance would be the full length of the rating variation or 70 percent (actual costs are limited to 1.35 x manual rates, and predicted costs are limited to 0.65 x manual rates in the error calculation, so the difference is 0.70 x manual rates).

The following grid, using the definitions just mentioned, presents the proportion of groups in each cell for traditional methods with a ± 35 percent rating variation, and one employee groups:

Traditional Method, 1 Employee Groups, ± 35%

ACTUAL \ PREDICTED			
	0%	15%	**6%**
	0%	**20%**	2%
	0%	54%	3%

When we ran regressions against historic loss ratios and manual rates for predicting costs, the weight for historic loss ratios was very low (about 15 percent). The previous table shows that the predicted costs do not go below the allowable rating variation because of this low credibility for historic costs.

The table below shows the results for the risk adjuster method (one employee groups, ± 35 percent rating variation). This method has a fairly high weight for the risk adjuster (about 80 percent). As can be seen by this table, 32 percent of the groups were predicted to have costs below the allowable rate variation.

Risk Adjuster Method, 1 Employee Groups, ± 35%

ACTUAL \ PREDICTED			
	2%	8%	**11%**
	3%	**13%**	7%
	27%	25%	5%

The risk adjuster puts 51 percent of the groups into the "correct" categories (highlighted on the diagonal), while the traditional method only puts 26 percent of the groups into the "correct" categories.

The following table shows the results for the risk adjuster method for groups of 50 employees:

Risk Adjuster Method, 50 Employee Groups, ± 35%

ACTUAL \ PREDICTED			
	0%	7%	**0%**
	0%	**91%**	0%
	0%	2%	0%

As this table shows, 91 percent of groups end up with costs within the allowable rate variation and are correctly predicted to be in this range. Further, the risk adjuster methodology does not estimate any groups outside of this range. The table for the traditional methodology is nearly identical.

These results show that traditional methods and renewal methods perform very similarly when rating limits and group size are introduced into the analysis.

Analysis on Carrier Data

As discussed earlier, we also modeled optimal methods on carrier data. This analysis addresses important limitations in the analysis based on the individualized data, including the following:

1. The individual analysis was performed by randomly creating groups from individual information as opposed to using actual small groups.
2. The full manual rating formula was available for the carrier's block of business.
3. The Standard Industry Code (SIC) information and rating variables were available. SIC rating variables estimate morbidity differences due to the industry of the group. Therefore, SIC rating adjustments can (typically) be used in addition to health status factors, essentially increasing the total allowable rate variation due to health status.
4. The actual turnover information was available. Because employees and members leave and enter employers, historic information (i.e., claims, diagnosis codes, etc.) is only available for a subset of the members being rated.

Our results for the carrier data show that methodologies that include risk adjusters lose much of their advantage under real-world conditions. Some of these real-world conditions could possibly be mitigated through process improvement (e.g., getting medical applications for new enrollees).

The absolute error is larger than that calculated in our individualized analysis because of turnover, and possibly adverse selection. The groups present in a block of business are only those that accepted a renewal rate. Therefore, they include groups who sought a new business quote from another carrier, and decided to accept the renewal quote.

These results show that traditional methods and renewal methods perform very similarly when rating limits and group size are introduced into the analysis.

While our analysis of the individual and carrier data considered the total health status factor variation allowed by states, we did not model the impact of state limits on the amount that the health status factor can vary from one year to the next (i.e., the health status factor cannot increase by more than 15 points). This additional constraint further limits rating action and dampens the predictive ability of any renewal method (i.e., you cannot move rates as much as your rating method predicts that you should).

We assumed that nine or 21 months (first renewal and second renewal respectively) of paid claims data would be available for both the traditional methods and risk adjuster methods when renewal rates were developed. If less data is available at the time renewal rates are developed, the differences between the two methods could change slightly. We would hypothesize that risk adjusters would lose less value because prescription drug databased risk adjusters perform well and prescription drug data completes more quickly than medical data.

The commercially available risk adjusters identified as top performers in the 2002 SOA study include ERGs, DxCGs, RxGroups and several others.[1] We used the ERG risk adjuster software in our analysis. This risk adjuster was identified as the most predictive in the SOA study. This study was focused on quantifying the predictive power of the commercially available risk adjusters. It was not intended to consider risk adjusters in the context of small group renewal rating.

(continued on page 23)

[1] *See the 2002 research study sponsored by the SOA, "A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations"*

Commercially available risk adjusters use member data and diagnosis information to assign each member into their demographic category and any relevant condition categories. Regression analysis can be used to best fit these category groupings to actual prospective costs. In this way, the specifics of the block of business can be reflected.

ERGs map each individual into their age/gender category and any of 120 condition categories. The condition categories include diabetes, heart failure and AIDS/HIV (an individual can be included in more than one condition category). ERGs also identify conditions where comorbidities are important. For example, there are separate condition categories for diabetes without comorbidities and diabetes with comorbidites.

In the study, we investigated whether the weights for different components should depend on the level of the factor and the group size. We concluded that the greater that the risk adjustment factors and/or loss ratio factors were, the greater the weight they should receive in the calculation.

Competitive Simulations

As noted earlier, renewal methods do not compete against other renewal methods. Instead, they compete against other carriers' new business methods. Therefore, simulation models should quantify the benefit of one method over another as the change in how those methods compete against new business methods.

New business methods are not straightforward to model, as new business rate setting often relies upon underwriter judgment. We had access to the uncapped new business health status factors assigned by one of our partner carriers. Using this information, we developed a Bayesian distribution for the new business HSFs. This distribution assigned the likelihood that an underwriter would assign various health status factors based on the actual outcome for the group (i.e. given that actual results were 150 percent greater than manual, what is the probability that the underwriter assigned a HSF of 0.81 to 0.90, 0.91 to 1.00, 1.01 to 1.10, etc.). This distribution resulted in a stochastic new business health status factor that we could compare against the health status factors assigned by various renewal methods.

Our initial simulation models indicate that renewal methods that use risk adjusters just slightly outperform renewal methods that do not use risk adjusters. The marginal value of improving renewal methods decreases as the predictive ability of your competitors' new business methods decrease. In other words, if you compete against carriers with very poor new business methods, you will realize less gain in profits by optimizing your renewal methods than if you compete against carriers with better new business methods. It follows that the best use of resources may be in improving new business predictability.

Conclusion

Optimal methods for small group rating depend on many variables, most significantly the goals of the company, size of the block of business, current competitive positioning, state regulations and available resources. One place to start improving your methods is to review how you are using the information and tools you currently have to ensure that you are using them optimally. Depending on your goals and the characteristics of your block of business, it may or may not make sense to invest in more sophisticated tools such as risk adjusters.