NCCI'S 2007 HAZARD GROUP MAPPING

by John P. Robertson

ABSTRACT

At the beginning of 2007, the NCCI implemented a new seven-hazard-group system, replacing the previous four-hazard-group system. This article describes the analysis that led to the assignment of classes to the new seven hazard groups.

1. INTRODUCTION

A *hazard group* is a collection of workers compensation classifications that have relatively similar expected excess loss factors over a broad range of limits. At the beginning of 2007, the National Council on Compensation Insurance (NCCI) implemented a new seven-hazard-group system, replacing the previous four-hazard-group system. The new hazard groups are not simply a subdivision of the previous four; they are a substantially different mapping of classes to hazard group. This article describes the analysis that led to the assignment of classes to the new seven hazard groups.

Under the previous NCCI four-hazard-group system, the bulk of WC exposure in NCCI states was concentrated in two hazard groups, as can be seen in Table 1.

Table 1
Distribution of Classes by Prior Hazard Group

NCCI Hazard Group	Number of Classes	Premium (billions)	% of Total Premium
I	38	$1.3	0.9%
II	428	$67.2	45.6%
III	318	$75.3	51.1%
IV	86	$3.6	2.5%

In our analysis, we considered whether a finer delineation would be possible, and what might be the optimal number of hazard groups. Apart from those considerations, hazard group assignments should be reviewed periodically because of changes over time in the insurance industry, technology, workplaces, and the evolution of the classification system and Workers Compensation infrastructure. The previous review had been done in 1993.

Prior Work

The prior NCCI hazard groups were developed by first identifying seven variables based on relative claim frequency, severity, and pure premium, which were thought to be indicative of excess loss potential [3]. These variables were the ratios of class to statewide weighted average:

1. serious to total claim frequency ratio
2. serious indemnity severity
3. serious medical severity
4. serious severity, including medical
5. serious to total indemnity pure premium ratio
6. serious medical to total medical pure premium ratio
7. serious pure premium to total pure premium ratio.

Because of the correlation between the severity variables and the pure premium variables the above seven variables were reduced to just the first, fourth, and last variables. A principal components[1] analysis was then done to determine the linear combination of these variables that maximized the proportion of the total variance explained by this linear combination. The linear combination so identified is called the first principal component and is the single variable that was used to sort classes into hazard groups. Determination of the optimal number of hazard groups was outside the scope of that study and so the number of hazard groups remained unchanged at four.

A very different approach was employed by the W.C. Insurance Rating Bureau of California (WCIRB) [4]. The WCIRB's objective was to group classes with similar loss distributions. They used two statistics to sort classes into hazard groups. The first statistic was the percentage of claims excess of $150,000. This statistic was thought of as a proxy for large loss potential. The second statistic measured the difference between the class loss distribution and the average loss distribution across all classes. The different hazard groups corresponded to different ranges of these two statistics. The results were checked by using cluster analysis on these two variables.

Overview

Our approach owes much to the prior work on the subject, yet it is quite distinct.

We sorted classes into hazard groups based on their excess ratios rather than proxy variables. As shown in Corro and Engl [5], a distribution is characterized by its excess ratios and so there is no loss of information in working with excess ratios rather than with the size of loss density or distribution function. Section 2 describes how we computed class-specific excess ratios.

Section 3 describes how we used cluster analysis to group classes with similar excess ratios, and how we determined that seven is the optimal number of hazard groups.

---

[1] See Johnson and Wichern [2] for a discussion of Principal Components.

2

In section 4 we compare the new hazard group assignments with the prior assignments.

Following the analytic determination of hazard groups, the tentative assignments were reviewed by several underwriters, and, based on this input, NCCI changed some assignments; we describe this in section 5.

Finally, section 6 recaps the key ideas of this study and the key features of the new assignments.

2. CLASS EXCESS RATIOS

Gillam [6] describes in detail the NCCI procedure for computing excess ratios by hazard group for individual states. An *excess ratio* or *excess loss factor* (*ELF*) is the ratio of losses excess of a given limit to total losses.[2] In the NCCI procedure, each ELF for a state and hazard group is a weighted average of ELFs by injury type specific to the state and hazard group. The ELFs for an injury type for a state and hazard group are derived from ELFs for the injury type in the state, adjusted to the estimated mean loss in the hazard group in the state. Injury types used by NCCI are Fatal, Permanent Total, Permanent Partial, Temporary Total, and Medical Only.

To put this in mathematical terms, let X_i be the random variable giving the amount of loss for injury type i in the state, and let X_i have density function $f_i(x)$ and mean μ_i. Let S_i be the normalized state excess ratio function for injury type i; that is

$$S_i(r) = E\left[\max\left(\frac{X_i}{\mu_i} - r,\, 0\right)\right] = \int_r^\infty (t - r)\, g_i(t)\, dt.$$

where $g_i(x) = \mu_i f_i(\mu_i x)$ is the density function of the normalized losses X_i / μ_i. For a hazard group j, the overall excess ratio $R_j(L)$ at limit L is

$$R_j(L) = \sum_i w_{i,j} S_i(L / \mu_{i,j}), \qquad\qquad (1)$$

where $w_{i,j}$ is the percentage of losses due to injury type i in hazard group j, so $\sum_i w_{i,j} = 1$,

and $\mu_{i,j}$ is the average cost per case for injury type i in hazard group j.

In the same way we can compute countrywide excess ratios for a given class by just knowing the weights and average costs per case by injury type for a class. We based these on the most recent five years of data as of April, 2005. This included claim counts and losses by injury type for the states where NCCI collects such data. Losses were developed, trended, and brought on-level to reflect changes in workers compensation

---

[2] Another common definition of excess loss factor is as the ratio of losses excess of a given limit to premium. We are concerned only with ratios of losses excess of given limits to total losses.

benefits. With some minor state exceptions, the same classes apply in all states. As such, we could estimate class excess ratios on a countrywide basis. Thus for each class, c, we had a vector

$$R_c = (R_c(L_1), R_c(L_2), ..., R_c(L_n))$$

of excess ratios at certain loss limits $L_1, L_2, ..., L_n$.

The credibility to assign to each class excess ratio vector is considered in the next subsection, and selection of the loss limits to use in the analysis is discussed in Section 3.

Credibility

In the prior review, the credibility given to a class was

$$z = \min\left(\frac{n}{n+k} \times 1.5, 1\right), \tag{2}$$

where n is the number of claims in the class and k is the average number of claims per class. This gives a class with the average number of claims 75% credibility and a class with at least twice the average number of claims full credibility. Figure 1 shows the credibility produced by this formula by size of class.
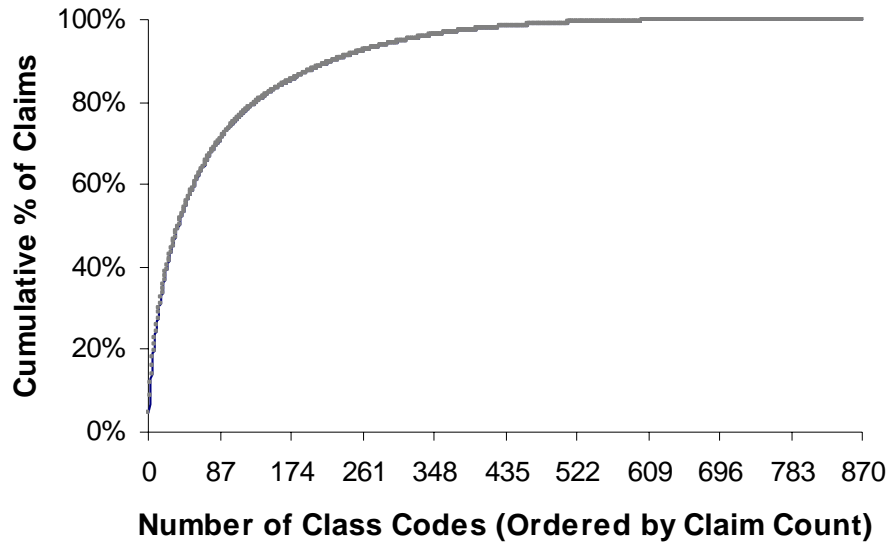
Figure 1
Class Code Credibility
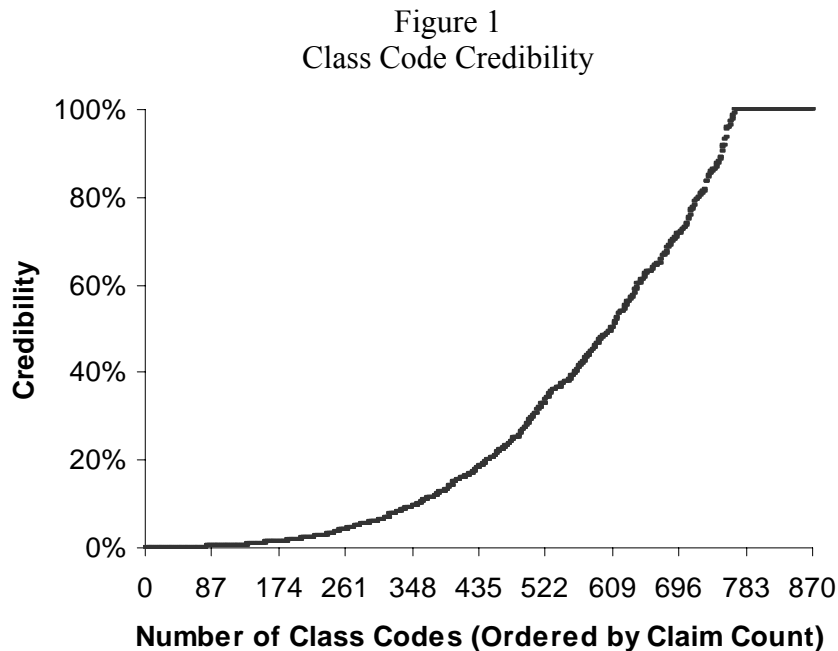
The fully credible classes have over 70% of the total premium as can be seen in Table 2.

Table 2
Distribution of Classes by Credibility

Credibility Range	Claims per Year	Number of Classes	% Premium
$0\% \leq z < 10\%$	0-237	355	1.2%
$10\% \leq z < 20\%$	238-511	89	1.3%
$20\% \leq z < 30\%$	512-831	61	1.6%
$30\% \leq z < 40\%$	832-1209	56	2.7%
$40\% \leq z < 50\%$	1210-1662	46	2.5%
$50\% \leq z < 60\%$	1663-2216	34	2.5%
$60\% \leq z < 70\%$	2217-2909	46	4.8%
$70\% \leq z < 80\%$	2910-3799	35	4.3%
$80\% \leq z < 90\%$	3800-4987	29	4.0%
$90\% \leq z < 100\%$	4988-6649	18	3.2%
$z = 100\%$	≥ 6650	101	71.8%
Total		870	100.0%

That the largest classes have a disproportionate share of the claims can be seen in Figure 2, where the classes with the greatest number of claims are to the left.

Figure 2
Distribution of Classes by Claim Count

Indeed the distribution of claims per class is very highly skewed, as can be seen in Figure 3.

Figure 3
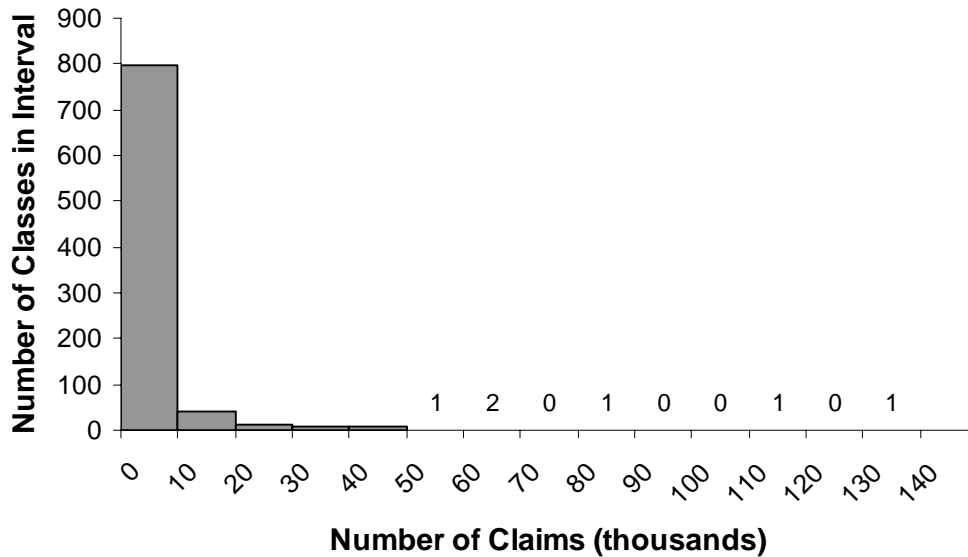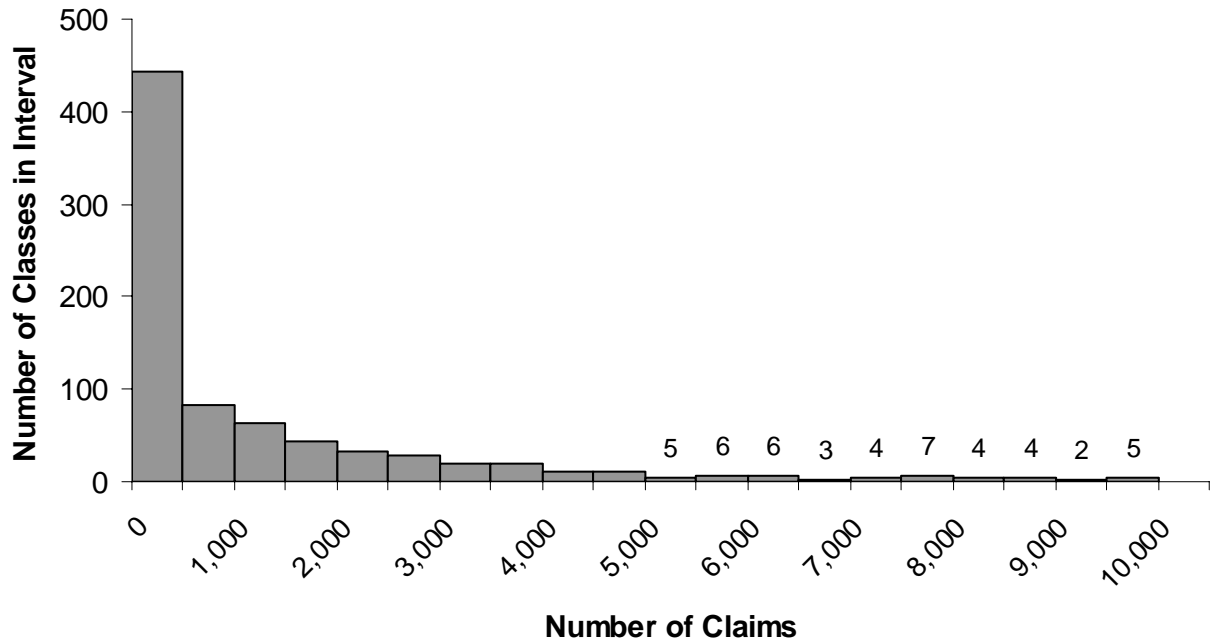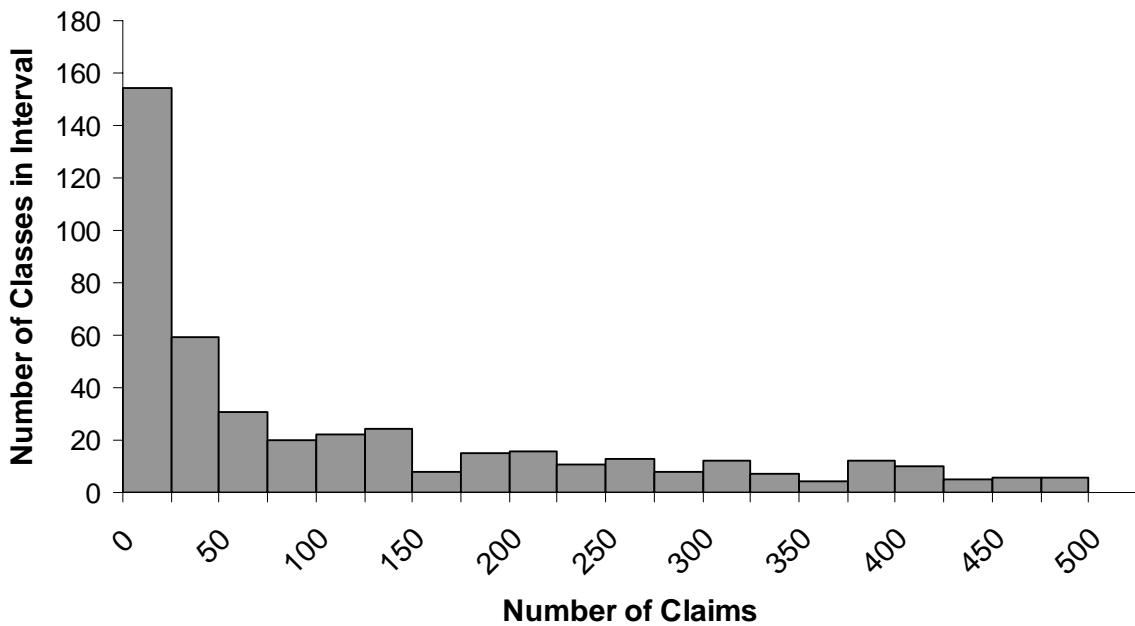Histogram of Number of Claims by Class

Figure 4 expands the first bar in Figure 3, and shows the persistency of the skewness.

Figure 4
Detail of Histogram of Number of Claims by Class

And Figure 5 further expands the first bar in Figure 4 revealing the same general pattern.

Figure 5
Detail of Histogram of Number of Claims by Class

The average number of claims per class is nearly ten times the median. We thus considered using the median rather than the mean for k in equation (2). This would have resulted in a very large increase in credibility as shown in the Figure 6.

Figure 6
Comparison of Credibility Formulas
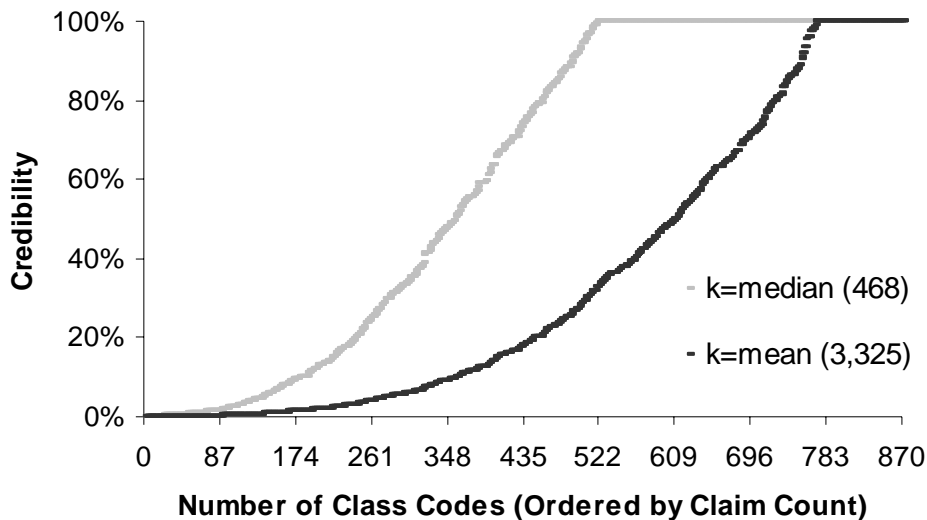
We considered several other variations on formula (2) as well. Because medical only claims have almost no impact on the ELFs at the published limits, we considered excluding all medical only claims. Taking that a step further, we looked at including only serious claims. We also considered taking k in formula (2) to be the mean number of claims over only those classes with some minimal number of claims.

In addition, we considered basing credibility on various square root rules. We considered a simple square root rule of the form

$$z = \sqrt{\frac{n}{384}},$$

where n is the number of claims in a class, and z is capped at 1. The full credibility standard of 384, given in Hossack, Pollard, and Zehnwirth [10, page 159], corresponds to a 95% chance of the actual number of claims within 10% of the expected number of claims. For the determination of ELFs, serious claims (Fatal, Permanent Total, and major Permanent Partial) are more important than non-serious claims, so we looked at the following variation on the square root rule above

8

$$z = \frac{N_F \sqrt{\dfrac{n_F}{384}} + N_M \sqrt{\dfrac{n_M}{384}} + N_m \sqrt{\dfrac{n_m}{384}}}{N_F + N_M + N_m},$$

where

> n_F = the number of fatal claims in the class
> N_F = the number of fatal claims in all classes
> n_M = the number of permanent total and major permanent partial claims in the class
> N_M = the number of permanent total and major permanent partial claims in all classes
> n_m = the number of minor permanent partial and temporary total claims in the class
> N_m = the number of minor permanent partial and temporary total claims in all classes.

We also considered varying the full credibility standard by injury type with the following credibility formula

$$z = \frac{N_s \sqrt{\dfrac{n_s}{175}} + \left(N - N_s \right) \sqrt{\dfrac{n - n_s}{384}}}{N}$$

where

> n_s = the number of serious claims in the class
> N_s = the number of serious claims in all classes
> n = the total number of claims in the class
> N = the total number of claims in all classes.

In the end, none of the alternatives considered seemed compelling enough to warrant a change and the results did not seem to depend heavily on the credibility formula; consequently we retained formula (2) for computing credibility.

For the complement of credibility we used the excess ratios corresponding to the current hazard group of the class. More precisely, for each class c we have a vector of excess ratios

$$R_c = (R_c(L_1), R_c(L_2),..., R_c(L_n))$$

and a credibility z. We also have a vector of excess ratios for the hazard group HG containing the class c (which can be determined, as above, as a loss weighted sum over vectors for classes in HG)

$$R_{HG} = (R_{HG}(L_1), R_{HG}(L_2),..., R_{HG}(L_n)).$$

We now associate to each class a credibility-weighted vector of excess ratios

$$zR_c + (1-z)R_{HG}.$$

It is these credibility-weighted vectors of excess ratios that we use in the cluster analysis described in the next section.

3. ANALYTIC DETERMINATION OF THE NEW HAZARD GROUPS

The fundamental analytic method used to determine the new hazard groups is Cluster Analysis. This is described in more detail below, but is a way to group classes with similar ELFs.

Selection of Loss Limits

The class excess ratio is a function of the loss limit, so it was necessary to select the attachment points to use in the analysis. We used limits of 100, 250, 500, 1000, and 5000, in thousands of dollars. Because excess ratios at different limits were highly correlated, five limits were thought to be sufficient. We considered using fewer limits but decided that it was better to use five limits to cover the range commonly used for retrospective rating.

We began by considering the 17 limits for which NCCI published excess loss factors before 2005. These limits, in thousands of dollars, were: 25, 30, 35, 40, 50, 75, 100, 125, 150, 175, 200, 250, 300, 500, 1000, 2000, and 5000. We modified this list by dropping $300,000 and adding $750,000. We reduced this to the five selected attachment points based primarily on two considerations:
- ELFs at any pair of excess limits are highly correlated across classes, especially when the ratio of the limits is close to 1.
- Limits below $100,000 are heavily represented in the list of 17 limits.

The correlations were computed using only the 162 classes with at least 75% credibility. Classes with small credibility have estimated ELFs close to those for the prior overall hazard group. Including the low-credibility classes would skew the correlations towards those of the overall hazard groups.

Even among the five selected limits, correlations between ELFs for pairs of limits are very high, as can be seen in Table 3.

Table 3
Correlations Among Excess Ratios at Selected Limits

Limit	100,000	250,000	500,000	1,000,000	5,000,000
100,000		0.992	0.973	0.935	0.824
250,000			0.994	0.969	0.879
500,000				0.990	0.925
1,000,000					0.968
5,000,000					

Each of the 12 limits not used has a correlation coefficient of at least 0.9882 with a limit that was used, as can be seen in Table 4.

Table 4
Correlations of ELFs for Pairs of Limits

Limits not Selected	Most Correlated Limit of the Five Selected	Correlation Coefficient
25,000	100,000	0.9882
30,000	100,000	0.9907
35,000	100,000	0.9926
40,000	100,000	0.9942
50,000	100,000	0.9965
75,000	100,000	0.9993
125,000	100,000	0.9996
150,000	100,000	0.9985
175,000	250,000	0.9987
200,000	250,000	0.9995
750,000	1,000,000	0.9982
2,000,000	1,000,000	0.9919

Although we ultimately used five limits, we experimented by clustering with different limits. We found that the hazard group assignments resulting from five limits were quite similar to those resulting from 17. When mapping the classes to seven hazard groups, only 68 out of 870 classes were assigned to different hazard groups and these accounted for just 5.5% of the total premium.

To see whether five limits were more than needed for the analysis, we tried clustering the classes using only a single limit. In one instance we used $100,000 and in another we used $1,000,000. Figures 7 and 8 compare those single limit assignments with clustering using the five-limit approach. In both cases, the results differed from the five-limit case, markedly so when $1,000,000 was used.

Figure 7
Clustering using $100,000 Limit Compared to Five Selected Limits
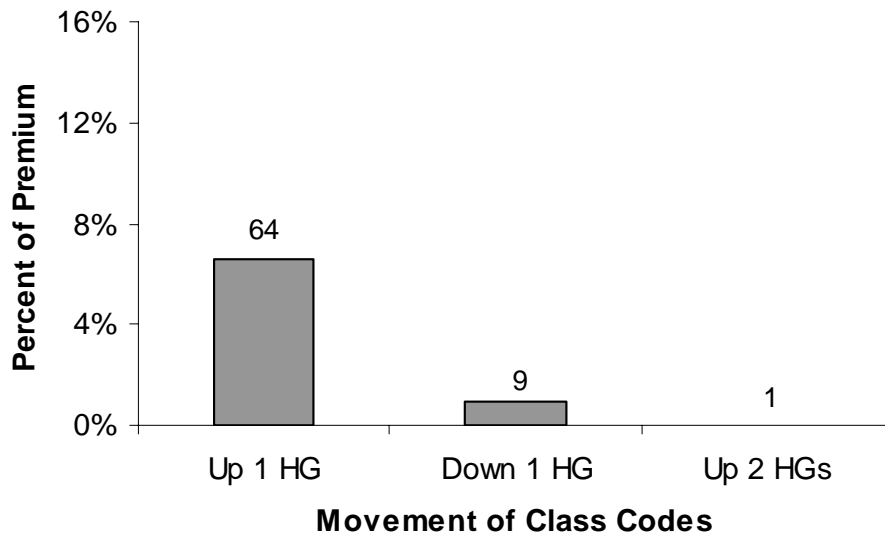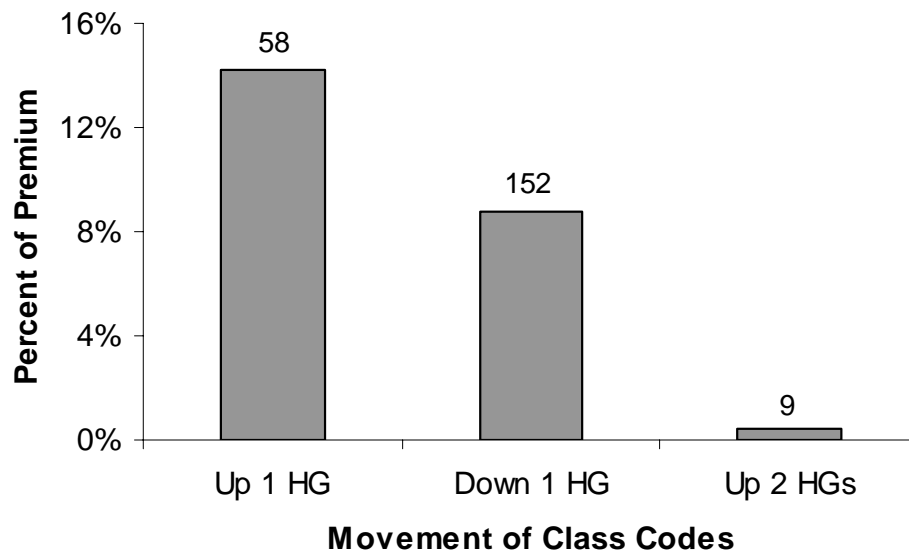(The number of classes that moved is shown above each bar.)

Figure 8
Clustering using $1,000,000 Limit Compared to Five Selected Limits
(The number of classes that moved is shown above each bar.)

This indicates that too much information is lost by dropping down to one limit. Retrospectively rated policies are purchased over a range of attachment points and no single limit captures the full variability in excess ratios.

We used principal components analysis to enhance the clustering investigation. The first two principal components of the five limits retained over 99% of the variation in the

data. While this might suggest that fewer limits could have been used, we decided to use five limits in order to cover the range of limits commonly used in retrospective rating. The distance between two classes in principal components space does not have the same simple interpretation as it does in excess ratio space. However principal components analysis allows one to project a five-dimensional plot onto two dimensions. Clustering using the five limits and plotting the resulting hazard group assignments using the first two principal components showed that the clusters were well separated and that outliers were easily identified. In our view, this confirmed the success of the five-dimensional clustering.

Metrics

The objective of assigning classes to hazard groups is to group classes with similar vectors of excess ratios. This raises the question of how to determine how similar or "close" two vectors are. The usual approach is to measure the distance between the vectors. If

$$x = (x_1, x_2, \ldots, x_n)$$

and

$$y = (y_1, y_2, \ldots, y_n)$$

are two vectors in $\mathbb{R}^n$, then the usual Euclidean, or L^2, distance between x and y is specified as

$$\left\| x - y \right\|_2 = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i \right)^2} \ .$$

This metric is used extensively in statistics and is what we used. In linear regression this metric penalizes large deviations. That is, one big deviation is seen to be worse than many small deviations.

There are many other metrics. Perhaps the second most common distance function is the L^1 metric which specifies

$$\left\| x - y \right\|_1 = \sum_{i=1}^{n} \left| x_i - y_i \right| .$$

Here a large deviation in one component gets no more weight than many small deviations. The intuitive rationale for using this metric is that it minimizes the relative error in estimating excess premium. If $R_c(L)$ is the hypothetically correct excess ratio at a limit of L for a class c and the premium on the policy is P then the excess premium is given by $P \cdot PLR \cdot R_c(L)$, where PLR denotes the permissible loss ratio. But in practice the class excess ratio is approximated by the hazard group excess ratio $R_{HG}(L)$. The relative error in estimating the excess premium is then

$$\frac{\left|P\cdot PLR\cdot R_{HG}(L)-P\cdot PLR\cdot R_c(L)\right|}{P}=PLR\cdot\left|R_{HG}(L)-R_c(L)\right|.$$

If we assume that each loss limit is equally likely to be chosen by the insured, then the expected relative error in estimating the excess premium is given by

$$\sum_{i=1}^{n}\frac{PLR}{n}\left|R_{HG}(L_i)-R_c(L_i)\right|=\frac{PLR}{n}\left\|R_{HG}-R_c\right\|_1,$$

which is proportional to the L^1 distance between the two excess ratio vectors.

Our analysis was not very sensitive to whether the L^1 or L^2 metric was used and we preferred the more traditional L^2 metric.

Standardization

When clustering variables measured in different units, standardization is typically applied to prevent a variable with large values from exerting undue influence on the results. Standardization ensures that each variable has a similar impact on the clusters. Duda and Hart [12] point out that standardization is appropriate when the spread of values in the data is due to normal random variation, however "it can be quite inappropriate if the spread is due to the presence of subclasses. Thus, this routine normalization may be less than helpful in the cases of greatest interest."

We considered two common approaches to standardization. The usual approach is to subtract the mean and divide by the standard deviation of each variable. For example, if $x_1, x_2, \ldots, x_n$ are the sample values of some random variable, with sample mean $\bar{x}$, and sample standard deviation s, then the standardized values are given by

$$z_i=\frac{x_i-\bar{x}}{s}.$$

An alternative standardization method depends on the range of observations. Under this approach we would take

$$z_i=\frac{x_i-\min x_i}{\max x_i-\min x_i}.$$

We conducted two cluster analysis trials in which we standardized according to the approaches described above. In each case we clustered the classes into seven hazard groups. Both trials resulted in hazard groups that were not very different from those produced without standardization.

Further, two issues were apparent with regard to standardizing in our particular analysis. First, excess ratios at different limits have a similar unit of measure, which is dollars of excess loss per dollar of total loss. That is, excess ratios share a common

denominator. Any attempt to standardize would have resulted in new variables without a common unit interpretation. Second, all excess ratios are between zero and one. Some standardization approaches would have resulted in standardized observations outside this range.

Another consideration is that excess ratios have a greater range at lower limits. Without standardization, the excess ratios at lower loss limits have more influence on the clusters than do those at higher limits. This is not undesirable because excess ratios at lower limits are based more on observed loss experience than on fitted loss distributions (see Corro and Engl [5]). Even on a national basis, there are few claims with reported losses above $5,000,000, but there are many more claims above $100,000. Greater confidence can be placed in the relative accuracy of excess ratios at lower limits because they are based on a greater volume of data.

In summary, the determination was made not to standardize because standardization would have eliminated the common denominator and it would have led to increased emphasis on higher limits. Our clustering algorithm used the L^2 metric and unstandardized credibility-weighted class excess ratios at the five selected loss limits: $100,000, $250,000, $500,000, $1,000,000 and $5,000,000. Premium weights were used to cluster the classes, as will be discussed in the next section.

Cluster Analysis

Given a set of *n* objects, the objective of cluster analysis is to group similar objects. In our case, we wanted to group classes with similar vectors of excess ratios, where similarity is determined by the L^2 metric. At this stage the number of clusters is taken as given. Typically partitions of the objects into 1, 2, 3, …, *n* clusters are considered. Non-hierarchical cluster analysis simply seeks the best partition for any given number of clusters. In hierarchical cluster analysis the partition with $k + 1$ clusters is related to the partition with *k* clusters in that one of the *k* clusters is simply subdivided to get the $k + 1$ element partition. Thus if two objects are in different clusters in the *k* cluster partition then they will be in different clusters in all partitions with more than *k* elements. This places a restriction on the clusters that can be sensible in some contexts. Our approach was non-hierarchical.

Optimality of k-means

The clustering technique we used is called *k*-means. For a given number, *k*, of clusters, *k*-means groups the classes into *k* hazard groups so as to minimize

$$\sum_{i=1}^{k} \sum_{c \in HG_i} \left\| R_c - \overline{R}_i \right\|_2^2 , \tag{3}$$

where the centroid

$$\overline{R}_i = \frac{1}{|HG_i|} \sum_{c \in HG_i} R_c$$

is the average excess ratio vector for the ith hazard group and $|HG_i|$ denotes the number of classes in hazard group i. Theoretically there is a difference between the hazard group excess ratio vector, R_{HG_i}, computed using (1), and the hazard group centroid, $\overline{R}_i$, but in practice this difference is very small.

Hazard groups determined by k-means have several desirable optimality properties. First, they maximize the following statistic

$$1 - \frac{\sum\limits_{i=1}^{k} \sum\limits_{c \in HG_i} \left\| R_c - \overline{R}_i \right\|_2^2}{\sum\limits_{c} \left\| R_c - \overline{R} \right\|_2^2}, \tag{4}$$

where

$$\overline{R} = \frac{1}{C} \sum_c R_c$$

is the overall average excess ratio vector, with $C = \sum |HG_i|$ being the total number of classes. Formula (4) is analogous to the R^2 statistic in linear regression. It gives the percentage of the total variation explained by the hazard groups.

A second way to evaluate hazard groups is based on the traditional concepts of within and between variance. We would like the hazard groups to be homogeneous and well separated. Thus we would like to minimize the within variance and maximize the between variance. We show now that using k-means accomplishes both.

Instead of considering a single excess ratio for each class, we have a vector of excess ratios. Thus we have not a single random variable corresponding to a single loss limit, but rather a random vector, with one random variable for each loss limit, from which we get a variance-covariance matrix. If X_i is the random variable for the excess ratio function at the ith loss limit, L_i, across classes c, then the observed values are the $R_c(L_i)$. The variance-covariance matrix of the random vector $X = (X_1, X_2, \ldots, X_n)$ is given by

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix},$$

where

$$\sigma_{ik} = E\big[(X_i - \mu_i)(X_k - \mu_k)\big]$$

is the covariance of X_i and X_k and $\mu_i = E[X_i]$. If we regard X as a 1 x n matrix then

$$\Sigma = E[(X - \mu)^T (X - \mu)],$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and $(X - \mu)^T$ is the transpose of $(X - \mu)$.

In practice the variance-covariance matrix is not known, but must be estimated from the data, i.e., the vectors

$$R_c = (R_c(L_1), R_c(L_2), \ldots, R_c(L_n)).$$

Let

$$\bar{x}_j = \frac{1}{C} \sum_c R_c(L_j),$$

where C is the total number of classes, and let

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n).$$

Then the sample covariance of the ELFs at L_i and L_k is

$$s_{ik} = \frac{1}{C} \sum_c (R_c(L_i) - \bar{x}_i)(R_c(L_k) - \bar{x}_k),$$

and the sample variance-covariance matrix is given by

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix} = \frac{1}{C} \sum_c (R_c - \bar{x})^T (R_c - \bar{x}).$$

One way to generalize the concept of variance to the multivariate context is to consider the trace of S

$$\text{trace}(S) = s_{11} + s_{22} + \cdots + s_{nn}.$$

This is just the sum of the sample variances of each variable and is called the total sample variance.

We let

$$T = CS = \sum_c (R_c - \bar{x})^T (R_c - \bar{x}).$$

The matrix T is proportional to the variance-covariance matrix for the whole data set. It is called the dispersion matrix, and is the matrix of sums of squares and cross products. We can proceed similarly within each hazard group and define

$$W_i = \sum_{c \in HG_i} \left(R_c - \bar{x}_i\right)^T \left(R_c - \bar{x}_i\right).$$

If we let

$$B_i = \left|HG_i\right| (\bar{x}_i - \bar{x})^T (\bar{x}_i - \bar{x}),$$

then it can be shown (see Späth [7]) that

$$\sum_{c \in HG_i} \left(R_c - \bar{x}\right)^T \left(R_c - \bar{x}\right) = B_i + W_i.$$

We then let

$$W = \sum_{i=1}^{k} W_i.$$

This is the pooled within group dispersion matrix. For the between variance we let

$$B = \sum_{i=1}^{k} B_i.$$

This is the weighted between group dispersion matrix. We then have

$$T = B + W.$$

This means, roughly that the total variance is the sum of the between variance and the within variance. Taking the trace we get

$$\text{trace}(T) = \text{trace}(B) + \text{trace}(W).$$

Thus the total sample variance is the sum of the between and within sample variance. Because trace(T) is constant, maximizing trace(B) is equivalent to minimizing trace(W), which is what k-means cluster analysis accomplishes.

Weighted k-means

As observed in section 2, some classes are much larger than others. To avoid letting the small classes have an undue influence on the analysis, we weighted each class by its premium. In simplest terms, this amounts to counting a class twice if it has twice as much premium as the smallest class. So instead of minimizing the expression in (3), we instead minimized

$$\sum_{i=1}^{k} \sum_{c \in HG_i} w_c \left\| R_c - \overline{R}_i \right\|_2^2 ,$$

where w_c is the percentage of the total premium in class c. We used the premium-weighted centroids as well, that is

$$\overline{R}_i = \frac{\sum_{c \in HG_i} w_c R_c}{\sum_{c \in HG_i} w_c} .$$

Optimal Number of Hazard Groups

So far, we have discussed the task of determining clusters when the number of clusters is given. We now address how to tell whether one number of clusters performs better than another, e.g., whether seven clusters works better than six or eight.

Various test statistics can be used to help determine the optimal number of clusters. The procedure is to compute the test statistic for each number of clusters under consideration and then identify the number of clusters at which the chosen statistic reaches an optimal value (either a minimum or a maximum, depending on the particular test statistic being used). Milligan and Cooper ([8], [9]) tested such procedures to determine which statistics were the most reliable.

Milligan and Cooper [8] performed a simulation to test 30 procedures. The simulated clusters were well separated from each other and they did not overlap. For each simulated data set, the true number of clusters was known, and they computed the number of clusters indicated by each method of determining the optimal number of clusters. The methods were ranked according to the number of times that they successfully indicated the correct number of clusters.

They noted that their simulation was idealized but that "It is hard to believe that a method that fails on the present data would perform better on less defined structures" ([8], page 161). Hence, although the hazard group data had both noise and overlap, it was useful to refer to Milligan and Cooper [8] to determine which methods to rule out.

In a later study, Cooper and Milligan [9] conducted tests that were more relevant to our application because random errors were added to the simulated data. That study found that the two best performing methods in the error-free scenario were also the best with errors ([9], page 319). The best performing method is due to Calinski and Harabasz. Milligan and Cooper ([8], page 163) define the Calinski and Harabasz statistic as

$$\frac{\text{trace}(B)/(k-1)}{\text{trace}(W)/(n-k)}$$

where n is the number of classes and k is the number of hazard groups, B is the between cluster sum of squares and cross product matrix, and W is the within cluster sum of squares and cross product matrix. A higher value of this statistic indicates better clusters because that corresponds to higher between clusters distances (the numerator) and lower within cluster distances (the denominator). This test is also known as the Pseudo-F test due to its resemblance to the F-test of regression analysis, often used to determine whether the explanatory variables as a group are statistically significant.

Another test that ranked high in the Milligan and Cooper testing was the Cubic Clustering Criterion (CCC). This test compares the amount of variance explained by a given set of clusters to that expected when clusters are formed at random based on data sampled from the multi-dimensional uniform distribution. If the amount of variance explained by the clusters is significantly higher than expected then a high value of the CCC statistic will result, indicating a high-performing set of clusters. An optimum number of clusters is identified when the test statistic reaches a maximum (Milligan and Cooper [8], page 164).

Milligan and Cooper [8] found that the Calinski and Harabasz test produced the correct number of clusters for 390 data sets out of 432. The CCC test produced the correct value 321 times. We could not use some of the other methods that ranked high because they were only applicable to hierarchical clustering, or for other reasons.

In a SAS Institute technical report, Sarle [11] noted that the CCC is less reliable when the data is elongated (i.e., variables are highly correlated). Excess ratios are correlated across limits, so we gave the CCC results less weight than the Calinski and Harabasz results.

We performed cluster analyses for four to nine hazard groups. There were four hazard groups in the prior NCCI system, and we saw no reason to consider any smaller number. Implementing ten or more hazard groups would be substantially more difficult than implementing nine or fewer, because having 10 or more requires an additional digit fro coding hazard groups. Testing up to nine was appropriate because the Workers Compensation Insurance Rating Bureau of California uses nine hazard groups [4].

In the first phase of our cluster analysis, we assigned classes and calculated the two test statistics for each number of groups under consideration. Figure 9 shows that the Calinski and Harabasz statistic indicated that the best number of hazard groups was seven. Figure 10 shows that the CCC statistic suggested nine hazard groups.

20

Figure 9
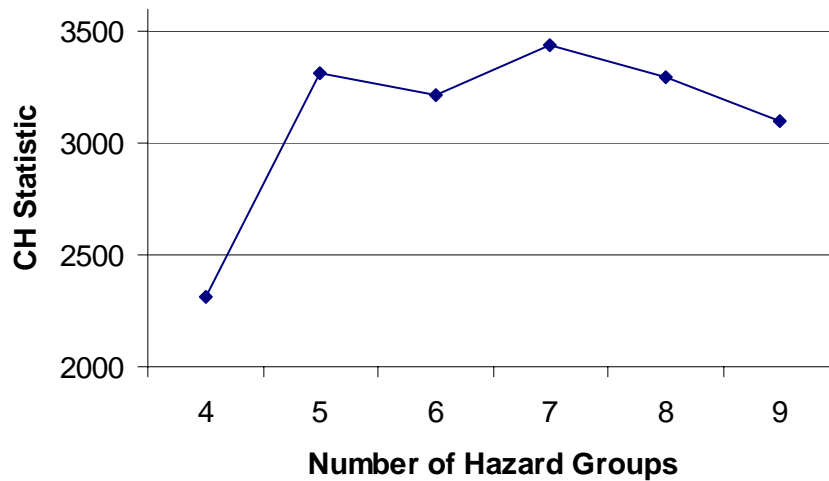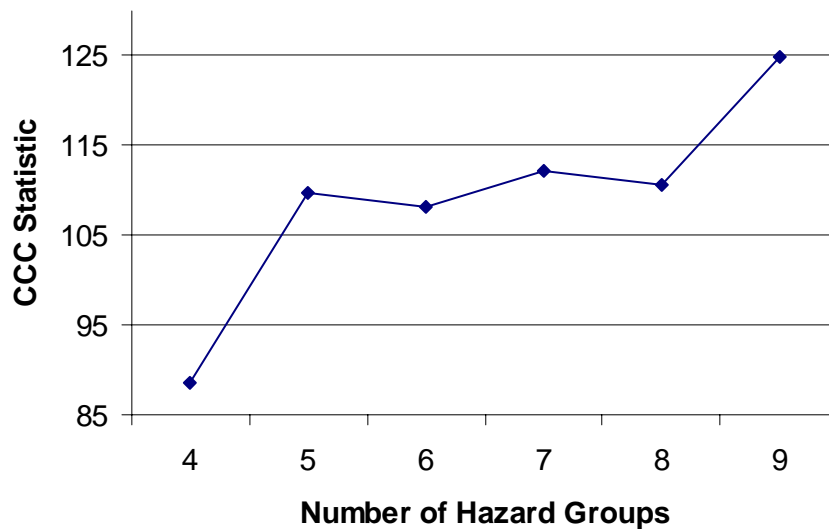Indicated Number of Hazard Groups
Calinski and Harabasz

Figure 10
Indicated Number of Hazard Groups
Cubic Clustering Criterion

But nine hazard groups produced crossover, meaning that the hazard group excess ratio for a higher hazard group was lower than that for a lower hazard group at some high loss limit. In our opinion, this suggested that more clusters were being used than could accurately be distinguished.

As can be seen in Table 2, most of the premium is concentrated in the largest classes with the highest credibility. We were concerned that the indicated number of hazard groups in the analysis above could have been distorted by the presence of hundreds of non-credible classes. In the second phase of our cluster analysis, we applied the tests to determine the optimal number of clusters using large classes only.

In one scenario, we applied the Calinski and Harabasz and CCC tests using only those classes with credibility greater than or equal to 50 percent. In a second scenario, we applied the tests using only fully credible classes. As shown in Figure 11, the indicated number of hazard groups was seven for both tests in both scenarios.

Figure 11
Statistics for Various Numbers of Hazard Groups
Only Classes with at Least 50 Percent Credibility
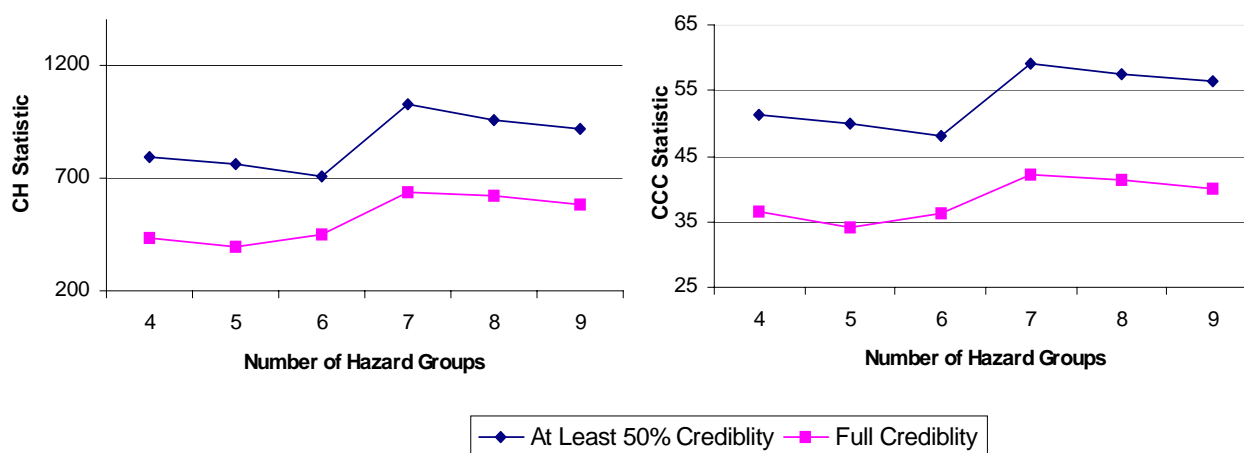
In summary, we used two test statistics in three scenarios for a total of six tests. Seven hazard groups was the indicated optimal number in five of these six tests. The exception was the scenario in which all classes were included, where the CCC test indicated that nine hazard groups were optimal. But there are four reasons why this exception received little emphasis:

- Milligan and Cooper ([8], [9]) found that the Calinski and Harabasz procedure outperformed the CCC procedure.
- The CCC procedure deserves less weight when correlation is present, which was the case in all of our scenarios.
- The selection of the optimal number of clusters ought to be driven by the large classes where most of the experience is concentrated. The large classes have the highest credibility and so the most confidence can be placed in their excess ratios.
- There is crossover in the nine hazard groups, and we don't think the data supports hazard groups with crossover.

22

We concluded that seven hazard groups were optimal. These are denoted A to G, with Hazard Group A having the smallest ELFs and Hazard Group G having the largest.

Alternate Mapping to Four Hazard Groups

We recognized that some insurers would not be able to adopt the seven hazard group system immediately because they needed additional time to make the necessary systems changes. Therefore we produced a four hazard group alternative to supplement the seven hazard group system. We chose to collapse the seven hazard groups into four by combining Hazard Groups A and B to form Hazard Group 1, combining C and D to form 2, combining E and F to form 3, and letting Hazard Group 4 be the same as G. Having an alternate mapping to four hazard groups simplifies comparisons between the prior and new mappings as well.

Prior to choosing this simple scheme we considered other alternatives. We tried using *k*-means cluster analysis to map the seven hazard group centroids into four. This approach resulted in a hazard group premium distribution that was not homogeneous enough. Another approach we considered was using cluster analysis to group the classes directly into four hazard groups. That approach yielded reasonable results, but it resulted in a non-hierarchical collapsing scheme, i.e., the seven hazard groups were not a result of subdividing the four hazard groups. The hierarchical collapsing scheme we chose has this feature, which allows users to know which of the four hazard groups a class is in based on knowing that class' assignment in the seven hazard group system.

The new four hazard group system is intended to be temporary. The four hazard group system is in place only to ensure that all carriers have sufficient time to make the transition to seven hazard groups.

4. COMPARISON OF NEW MAPPING WITH OLD

Distribution of Classes and Premium

The bulk of the exposure was concentrated in two of the hazard groups prior to our review. Hazard Groups I and IV contained a small percentage of the total premium. Hazard Groups II and III, on the other hand, contained 97 percent of the total premium (see Table 1). We knew that a more homogeneous distribution of premium by hazard group would improve pricing accuracy. When discussing the new hazard groups in this section we will focus on the mapping that resulted directly from the statistical analysis. Later on, as will be discussed in the underwriting review subsection, numerous classes were reassigned among the groups based on feedback gathered in our survey of underwriting experts. These changes are not reflected in Figures 12 to 20.

Figures 12 and 13 compare the prior mapping to the collapsed new mapping based on the distribution of classes and premium.

Figure 12
Prior Mapping vs. Collapsed New Mapping
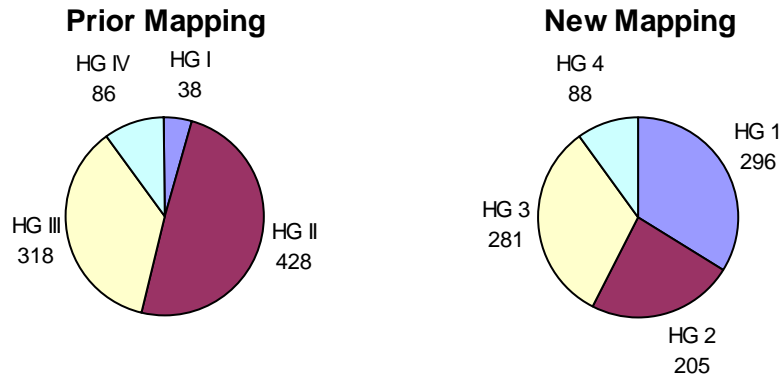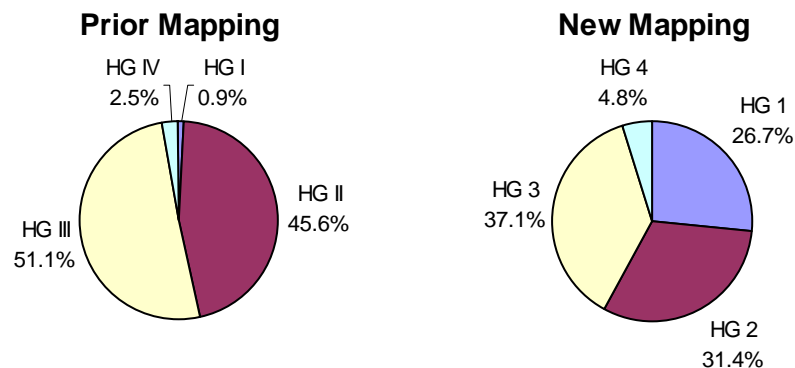Number of Classes per Hazard Group

Prior Mapping

HG IV 86 • HG I 38 • HG III 318 • HG II 428

New Mapping

HG 4 88 • HG 1 296 • HG 3 281 • HG 2 205

Figure 13
Prior Mapping vs. Collapsed New Mapping
Percent of Premium by Hazard Group

Prior Mapping

HG IV 2.5% • HG I 0.9% • HG III 51.1% • HG II 45.6%

New Mapping

HG 4 4.8% • HG 1 26.7% • HG 3 37.1% • HG 2 31.4%

Hazard Group 1 has a large number of classes and a substantial portion of total premium in contrast to Hazard Group I. Hazard Groups 2 and 3 have become slightly smaller than before although they are still large. In the prior mapping Hazard Groups II and III each had over 45 percent of the premium, but in the new mapping, none of the four groups has as much as 40 percent. This refinement allows for improved homogeneity of classes within each hazard group. Hazard Group 4 has retained a similar number of classes but it has more premium than Group IV.

Figure 14 shows that most of the classes and premium remained in the same hazard group when assigned to the new four Hazard Groups.

Figure 14
Comparison of Old with New Assignment to Four Hazard Groups
(The number of classes that moved is shown above each bar.)
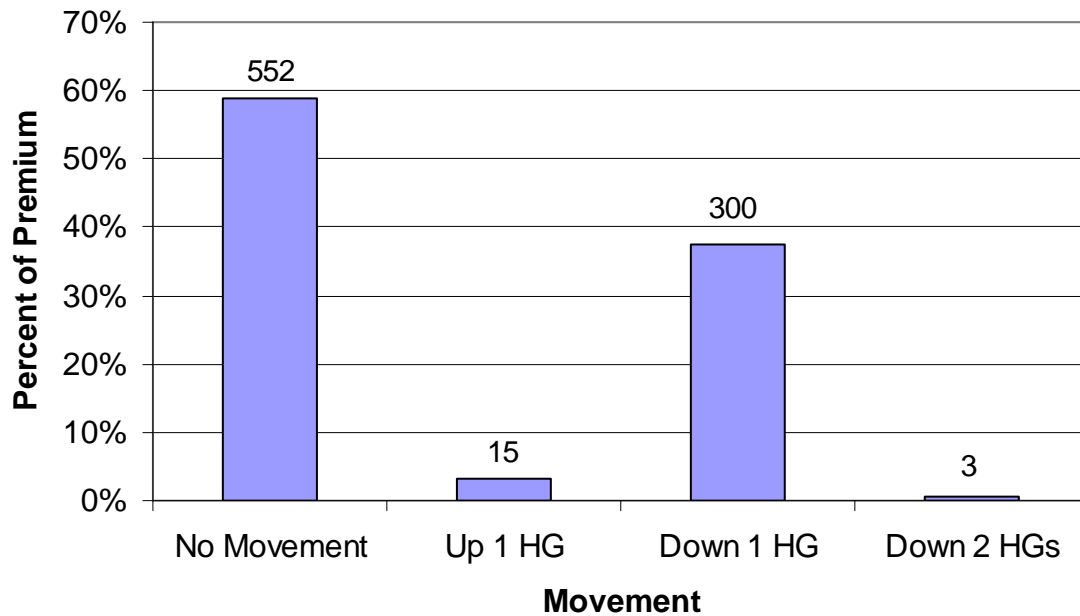
Among those classes that did move, the great majority (300 classes and 37 percent of the premium) moved down one hazard group. Most of this movement was from Hazard Group II to 1. The movements of classes and premium are detailed in Table 5.

3/17/09

Table 5
Comparison of Distributions of Classes Between
Prior and New Hazard Group Assignments

Prior Mapping

Hazard Group	I	II	III	IV	Total
Number of Classes	38	428	318	86	870
% Premium	0.9%	45.6%	51.1%	2.5%	100%

New Mapping / **Hazard Group**

	I	II	III	IV	Total
1	38	255	3	0	296
	0.9%	25.4%	0.5%	0.0%	26.7%
2	0	164	41	0	205
	0.0%	19.6%	11.8%	0.0%	31.4%
3	0	9	268	4	281
	0.0%	0.6%	36.3%	0.2%	37.1%
4	0	0	6	82	88
	0.0%	0.0%	2.6%	2.2%	4.8%

The table can be read vertically. For instance, among the 428 classes in Hazard Group II, 255 were mapped into Hazard Group 1, 164 into Hazard Group 2, nine into Hazard Group 3, and none into Hazard Group 4. The 255 classes that moved from Hazard Group II into Hazard Group 1 comprised 25.4% of the total premium. A significant number of classes and amount of premium moved from Hazard Group III to 2. Three classes moved from III to 1. Just 15 classes moved up by one hazard group, making up three percent of the premium. Hazard Group 1 is so large primarily because of classes that entered it from Hazard Group II. Hazard Group 2 is quite different than Hazard Group II because many of the classes in 2 originated in III and many of the classes that were in II have moved into 1.

The new seven hazard group assignment has a fairly homogenous distribution of classes and premium, as shown in Figure 15.

© Copyright 2009 National Council on Compensation Insurance, Inc. All Rights Reserved.   26

Figure 15
Number of Classes and Percent of Premium in Each Hazard Group

Number of Classes per HG **Percent of Premium per HG**
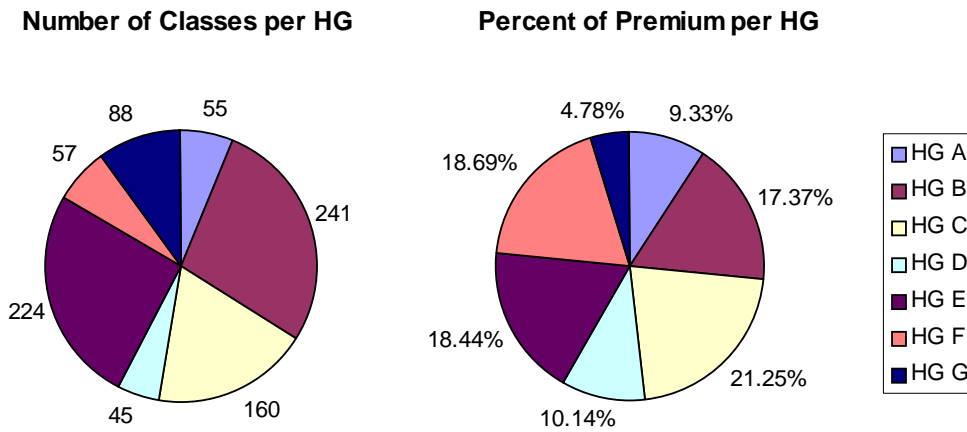
This is a marked improvement over the prior mapping. In terms of premium, Hazard Group A is 11 times larger than Hazard Group I was. Hazard Group G is twice as large as Hazard Group IV was.

Table 6 shows the distribution of classes to hazard groups based on their level of credibility. Overall there were 162 classes with at least 75 percent credibility and 708 classes with lower credibility. Generally, within each hazard group most of the premium is due to highly credible classes but most of the classes have lower credibility. Hazard Groups D and G are exceptions. Hazard Group D has nearly equal numbers of high and low-credibility classes. In Hazard Group G, high and low-credibility classes have similar premium percentages.

Table 6
Number of Classes with Given Credibility by Hazard Group

Hazard Group	162 Classes with Credibility $\geq$ 75%		708 Classes with Credibility < 75%	
	Number of Classes	% Premium	Number of Classes	% Premium
A	18	8.4%	37	0.9%
B	40	14.5%	201	2.8%
C	41	17.6%	119	3.6%
D	22	8.9%	23	1.3%
E	22	14.0%	202	4.4%
F	15	15.0%	42	3.7%
G	4	2.4%	84	2.4%
Total	162	80.9%	708	19.1%

Although Hazard Groups B and E have far more classes than the other hazard groups, they do not have far more premium. The reason that they have the most classes with credibility less than 75 percent is that the complement of credibility is the prior hazard group excess ratio. For instance, the excess ratio of Hazard Group III at $100,000 was 0.451 which is close to the excess ratio of Hazard Group E. Given a small class in Hazard Group III, the credibility-weighted excess ratio was likely to be close to the excess ratio of Hazard Group E.

Range of Excess Ratios

In Figure 16 each horizontal bar represents the range of credibility-weighted excess ratios within a particular hazard group. The vertical line within each bar represents the overall excess ratio for the hazard group.
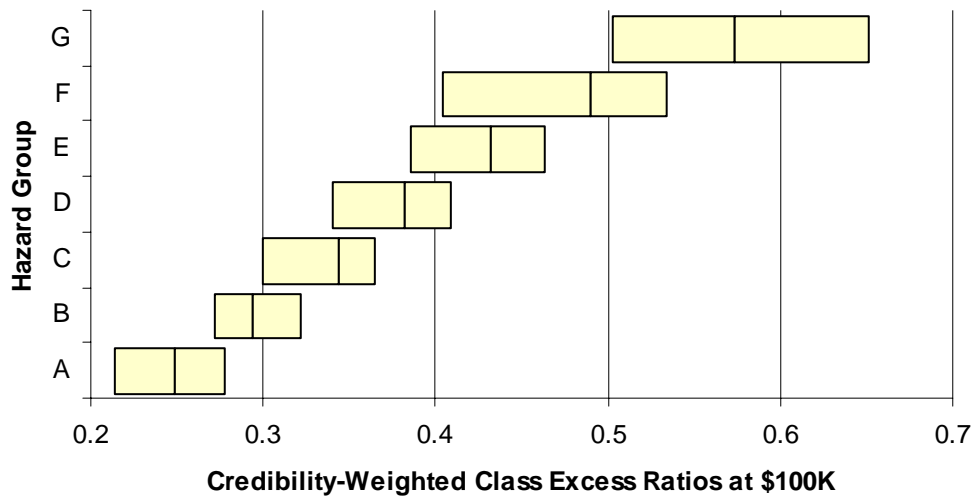
Figure 16
Prior Mapping Excess Ratio Ranges at $100K
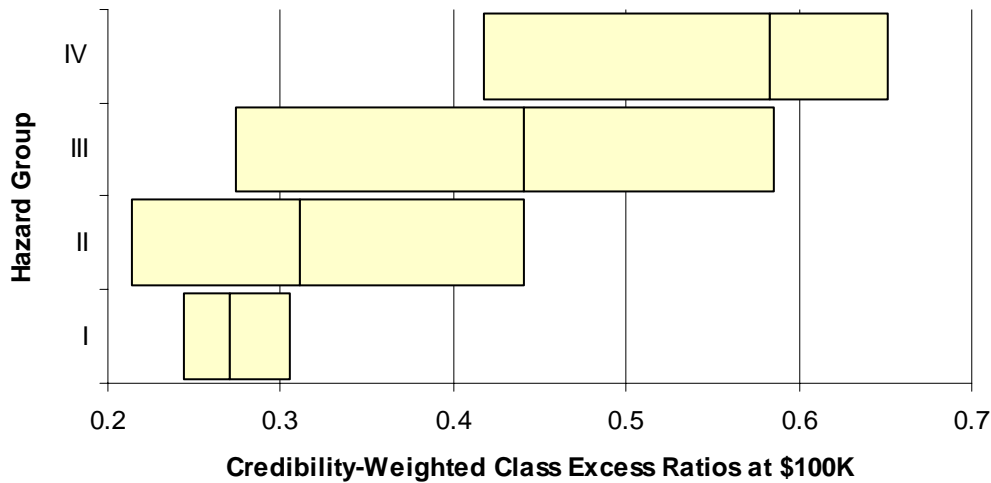
Among the classes in Hazard Group I, the excess ratios at $100,000 ranged from 0.254 to 0.315. In Hazard Group II, the excess ratios at $100,000 ranged from 0.223 to 0.451. Thus the range of Hazard Group I excess ratios was contained within that of Hazard Group II, indicating that Hazard Groups I and II were not as well-separated as might be desired. The same behavior was observed at $1,000,000 as well.

As shown in Figure 17, *k*-means clustering resulted in well-separated hazard groups. Because five dimensions were used, we could not avoid overlap in each dimension, but the excess ratio distribution is a noticeable improvement over the prior mapping.

29

Figure 17
New Mapping Excess Ratio Ranges at $100K

The new mapping also shows a well-separated excess ratio distribution at $1,000,000 as shown in the Figure 18.

Figure 18
New Mapping Excess Ratio Ranges at $1M
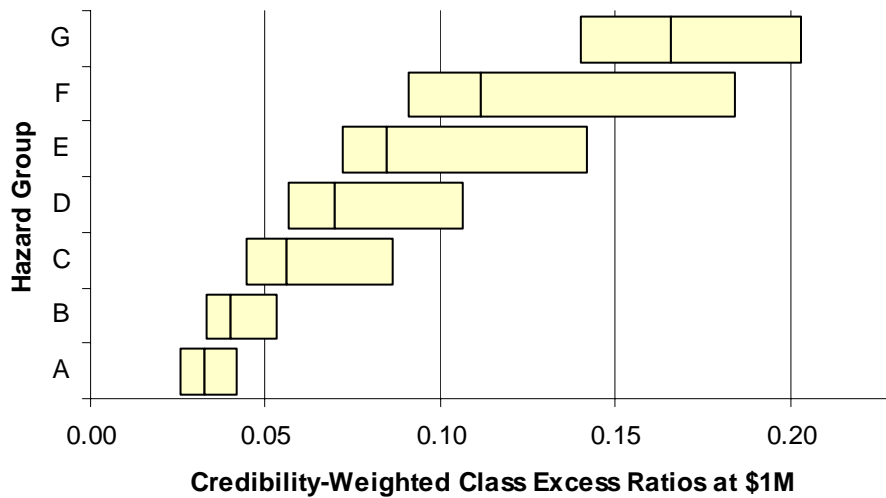
Most of the exposure is concentrated in the largest classes, and so the hazard group excess ratios are highly sensitive to the placement of large classes. In Figures 16- 18, the range of excess ratios for each hazard group is calculated using all of the classes in that hazard group.

Figures 19 and 20 show that if ranges are computed using only those classes with at least 75 percent credibility, then the separation of hazard groups by excess ratios is quite strong at both $100,000 and $1,000,000.

Figure 19
New Mapping Excess Ratio Ranges at $100K
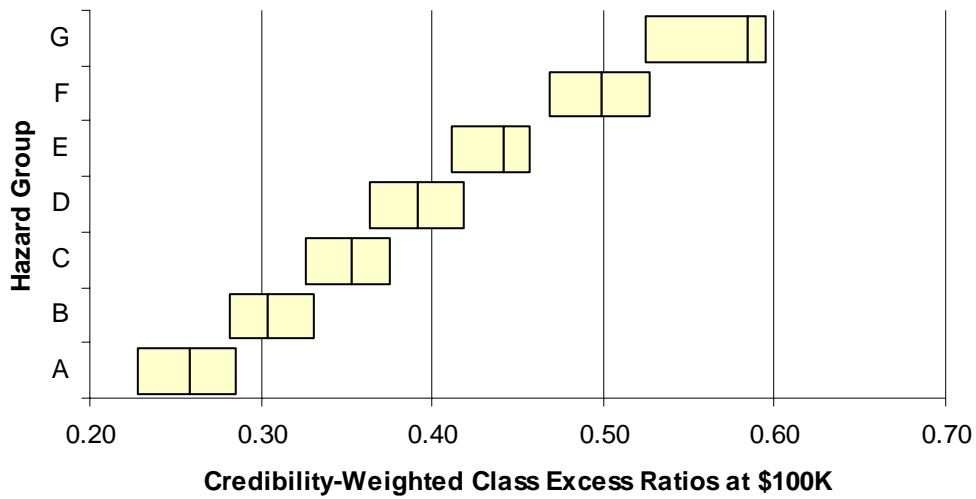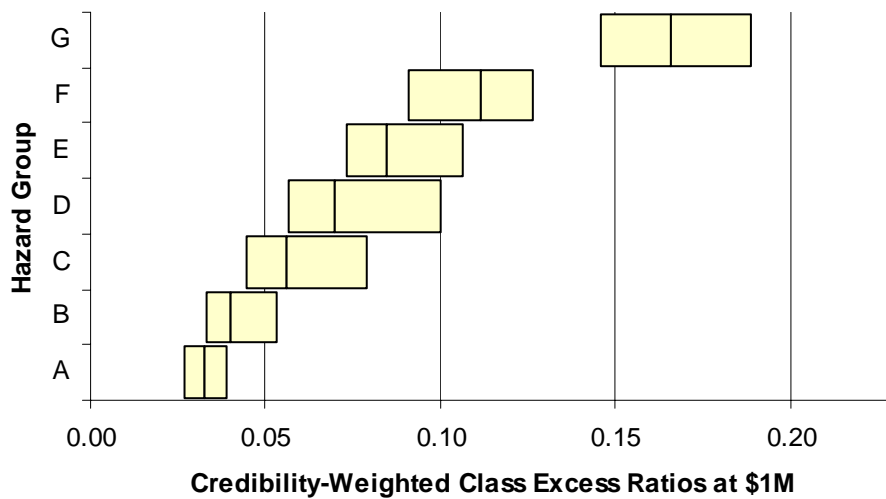Classes with at least 75% Credibility

Figure 20
New Mapping Excess Ratio Ranges at $1M
Classes with at least 75% Credibility

5. UNDERWRITING REVIEW

After completing the cluster analysis, we conducted a survey of underwriters to solicit their comments on the proposed new mapping. The survey was sent to all members of NCCI's Underwriting Advisory List (UAL), and included the draft mapping that resulted from the analytic determination of the hazard groups. The survey asked the underwriters to judge the hazardousness of each class based on the likelihood that a given claim would be a serious claim. We also pointed out that if the mix of operations in two classes was very similar then the two classes should probably be in the same hazard group.

Members of the UAL recommended changes in the hazard group assignment for a third of the classes. We also received feedback from two underwriters on staff at NCCI. After the survey comments were compiled, a team consisting of NCCI actuaries and underwriters reviewed the comments from UAL members and decided on the final assignment for each class. When deciding whether to reassign a class, we considered whether the feedback on that class was consistent. We considered the credibility of each class and placed more weight on the cluster analysis results for those classes with a large volume of loss experience. For each class we compared the excess ratios to the overall hazard group excess ratios and identified the nearest two hazard groups.

Class 0030 illustrates the process used at NCCI to decide on the hazard group for each class. This is the class for employees in the sugar cane plantation industry and is only applicable in a small number of states. This class
- had 12% credibility,
- was in Hazard Group III under the prior mapping, and
- was assigned to Hazard Group E under the cluster analysis.

An underwriter pointed out that Class 0030 has operations similar to Class 2021, which is for employees who work at sugar cane refining. Class 2021
- applies nationally,
- had 31% credibility,
- was in Hazard Group II under the prior mapping,
- was assigned to Hazard Group C under the cluster analysis, and
- prior to credibility weighting had excess ratios close to the overall excess ratios for Hazard Group D.

Credibility weighting had reduced Class 2021's excess ratios so that they were between the overall excess ratios of Hazard Groups C and D, because the prior assignment of Class 2021 had been to Hazard Group II.

We concluded that Hazard Group D was the best choice for 2021 based on its excess ratios prior to credibility weighting and its mix of operations. We determined that 0030 should be assigned to the same hazard group as 2021, so we also assigned Class 0030 to Hazard Group D.

Underwriters made several other types of comment besides those comparing one class to another. For instance, they commented on the degree to which employees in a given

32

class are prone to risk from automobile accidents. They commented on the extent to which heavy machinery is used in various occupations and how much exposure there is to dangerous substances.

Figure 21 displays the movements of premium and classes during the underwriting review under the collapsed new mapping. It shows that the overall effect of the underwriting review was to move a significant number of classes up to a higher hazard group.

Figure 21
Percent of Premium That Moved During the Underwriting Review
(The number of classes that moved is shown above each bar.)
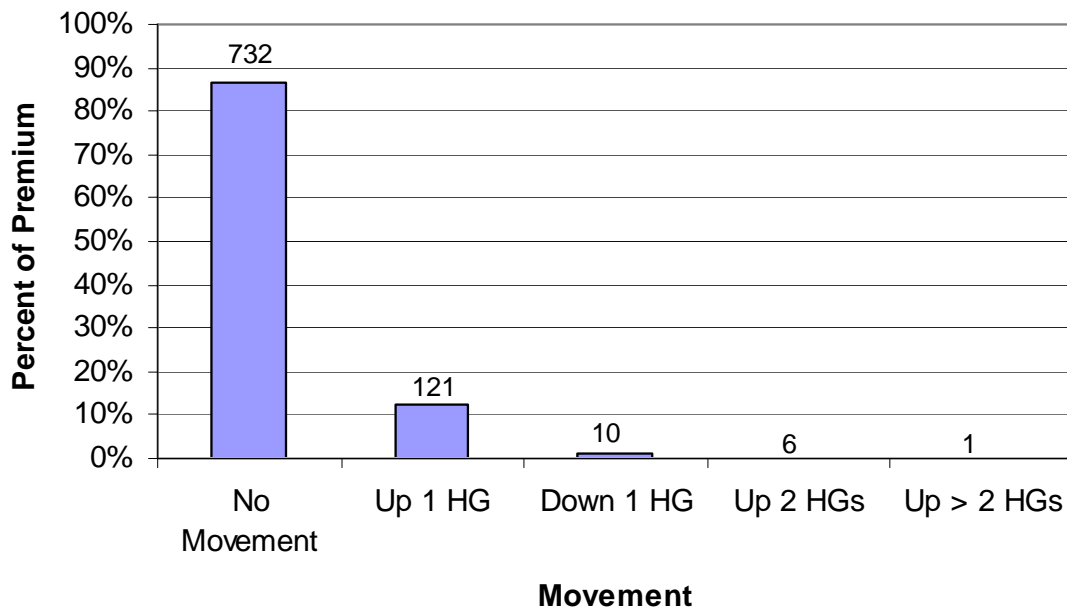
The majority of the classes that moved up one hazard group, 78 of them, moved from Hazard Group 1 to 2, while 20 classes moved from Hazard Group 2 to 3, and 23 classes moved from Hazard Group 3 to 4.

6. CONCLUSION

Our approach to remapping the hazard groups was founded on three key ideas:

1. Computing excess ratios by class
   The data is too sparse to directly estimate excess ratios by both class and state. But countrywide excess ratios can be computed by class in the same way that hazard group excess ratios are computed. This does not require separate loss distributions for each class. The existing loss distributions by injury type can be used along with the usual scale assumption. Thus all that is needed is average

costs per case by injury type and injury type weights for each class.

2. Sorting classes based on excess ratios
   Rather than using indirect variables to capture the amorphous concept of "excess loss potential," we used excess ratios directly because hazard groups are indeed used to separate classes based on excess ratios. Because a loss distribution is in fact characterized by its excess loss function, this approach involves no loss of information. By sorting classes based on excess ratios we achieve the goal of sorting classes based on their loss distributions as well.

3. Cluster Analysis
   Problems involving sorting objects into groups are not unique to actuarial science. We were thus able to make use of a large statistical literature on cluster analysis. This provided an objective criterion for determining the hazard groups as well as the optimal number of hazard groups. Our approach to determining the seven hazard groups was non-hierarchical because we wanted the best seven group partition and because hypothetical partitions into six hazard groups are not relevant in this context.

As a result of our analysis the number of NCCI hazard groups was increased from four to seven. The distribution of both premium and classes is much more even in the new hazard groups. The highest hazard group is still relatively small. The new seven hazard groups collapse naturally and hierarchically into four hazard groups. Comparing the new four hazard groups with the old, over two-thirds of the classes, with nearly 60% of the premium, did not move at all. This stability was largely a result of the fact that we used the old hazard group as a complement of credibility and there were a large number of classes with very little premium. Of the classes that did move, the overwhelming majority moved down one hazard group.

The new mapping was filed in mid-2006 to be effective with the first rate or loss cost filing in each state on or after January 1, 2007. The filing (Item Filing B-1403) was approved prior to the end of 2006 in all states in which NCCI files rates or loss costs.

ACKNOWLEDGMENTS

REFERENCES

[1] Mahler, Howard C., "Workers Compensation Excess Ratios: An Alternative Method of Estimation," *PCAS* LXXXV, 1998, pp. 132-156, http://www.casact.org/pubs/proceed/proceed98/980132.pdf.
[2] Johnson, Richard A. and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, 2002.

[3] National Council on Compensation Insurance, "Revised Hazard Group Assignments," Actuarial Committee Agenda, Item ACT-92-57, April, 20, 1993.

[4] W.C. Insurance Rating Bureau of California, "Retrospective Rating Methodologies—Hazard Group Assignment," Actuarial Committee Agenda, Item AC01-05-02, May 30, 2001.

[5] Corro, Dan and Greg Engl, "The 2004 NCCI Excess Loss Factors," Casualty Actuarial Society *Forum*, Fall 2006, pp. 513-571, http://www.casact.org/pubs/forum/06fforum/517.pdf.

[6] Gillam, Willam R., "Retrospective Rating: Excess Loss Factors," *PCAS* LXXVIII, 1991, pp. 1-40, http://www.casact.org/pubs/proceed/proceed91/91001.pdf.

[7] Späth, Helmuth, *Cluster Dissection: theory, FORTRAN programs, examples*, Halsted Press, 1985.

[8] Milligan, G. W. and M. C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika,* 50, 1985, pp. 159-179.

[9] Cooper, M. C. and G. W. Milligan, "The Effect of Measurement Error on Determining the Number of Clusters in Cluster Analysis," in Gaul, W., and M. Schader (eds.), *Data, Expert Knowledge, and Decisions: An Interdisciplinary Approach With Emphasis on Marketing Applications*, Berlin: Springer-Verlag, 1988, pp. 319-328.

[10] Hossack, I. B., Pollard, J. H., and Zehnwirth, B., *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, 1983.

[11] Sarle, W. S., *Cubic Clustering Criterion* (Technical Report A-108), Cary, NC: SAS Institute, 1983.

[12] Duda, R. O. and P. E. Hart, *Pattern Classification and Scene Analysis*, New York, NY: Wiley, 1973.