RECORD OF SOCIETY OF ACTUARIES
1992 VOL. 18 NO. 1B

MEDIUM- AND LARGE-GROUP MEDICAL ISSUES (BASIC)

Instructors: CHARLES S. FUHRER
ARNOLD F. SHAPIRO

- Experience-rating and credibility in group health insurance

MR. CHARLES S. FUHRER: If we covered all the topics listed in the program, this session would be a three-day seminar. So we chose to focus on one aspect and that is the prospective experience-rating and credibility in group health insurance. Arnold Shapiro is a professor at Penn State, which I was told to mention is the university with the football team. He heads up the actuarial program there, and he's going to give us some background on experience-rating and credibility.

MR. ARNOLD F. SHAPIRO: Since Chuck is going to deal with some specific applications, I will provide some background and some food for thought.

One of the questions actuaries have had to address is the extent to which credibility theory could be applied in their everyday work. In the old days, when what we call "Bayesian analysis" was referred to as the "inverse probability rule," many actuaries were concerned that the concept really had limited application in the insurance area.[1] We have come a long way since then, and now most would generally agree with the statement of Donald C. Baillie that " . . . actuaries are all basically Bayesian, whether they know it or not" [*TSA* XIX (1967): 122].

Of course, it is one thing to have a sense of credibility theory, and it is quite another to get the idea across to someone who has never been exposed to the notion. To this end, there are numerous examples in the literature. Often, the examples involve shooting at some sort of target, and some of those examples are referred to below. Another approach, however, which might be used to emphasize the importance of the weights associated with past and current information, is as follows.

Imagine, if you will, that there is a goblet spinning in space, as shown in Chart 1. The problem is to determine how it will land when it hits the ground. Will it land straight up? Will it land upside down? Will it land on its side?

Well, we have some sense of the likely outcome. The first question, of course, is whether the goblet is biased in some way. Is it weighted at the bottom? Assume it is not biased. Consequently, one would expect that when it hits the ground, it would flop over on its side. Of course, that is based on our prior expectation. If, just before it hits the ground, it is caught, held upright, and then dropped from a distance of just a couple of goblet heights, or less, our expectation would immediately change.

---

[1] A prime example of this was the following assertion which appeared in the *Trans. Actuarial Soc. Edimb.* (p. 421), in 1891, "Both from the point of view of practical common-sense, and the point of view of logic, the so-called laws of Inverse Probability are a useless appendage to the first principles of probability, if indeed they be not a flat contradiction of those very principles.... The laws of Inverse Probability being dead, should be decently buried out of sight, and not embalmed in text-books and examination papers." The author was an actuarial professor named G. Chrystal.

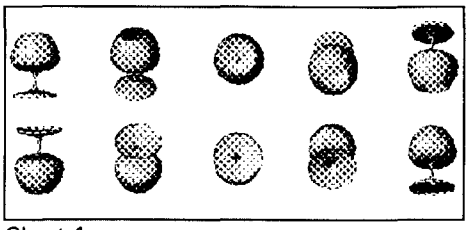

Chart 1

The implication is that the most recent information can be extremely relevant in certain situations, even if it is based on far less data than our prior expectations. In particular, where insured lives are involved, even the experience of small groups can be material when experience-rating.

The origin for our analysis was the article by Thomas Bayes, "An Essay towards Solving a Problem in the Doctrine of Chances," which appeared in the *Philosophical Transactions* in 1763. It was Dr. Richard Price (Northampton Tables) who actually published the article two years after Bayes's death.[2]

The early impetus for much of the work in the U.S., however, was embodied in the article of Albert W. Whitney, "The Theory of Experience-Rating" (*PCAS*, 1918), which dealt with property and casualty issues.

One of the things he emphasized was the trade-off between the class and the risk, and the relative weight to associate with each. Do you weight the class more, do you rate the risk more, and what is the relationship between the two?

In the group insurance area, we tend to regard the Bayesian approach to experience-rating, and the associated models as a part of the current thinking. In fact, however, the current approach is quite similar to that taken by Whitney. His conceptualization of the credibility factor, Z, took the form $Z = \frac{x - P}{p - P} = \frac{allowable\ \ deviation}{indicated\ \ deviation}$ where x is the adjusted rate for the risk (the allowable rate), P is the average rate for the risk (the overall mean), and p is the rate indicated by the risk's own experience. Note that this formulation, rewritten in the form $x = Zp + (1-Z)P$ is the same one that we currently use. That is, the adjusted rate is equal to the credibility factor times the personal experience plus 1 minus the credibility factor times the population mean.

In the years that followed, numerous studies implemented this formula. Representative of these was the paper by Robert A. Bailey[3] and Leroy Simon, "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car" (*PCAS*,

---

[2]Bayes was a minister who was interested in mathematical problems. He had not produced the theorem known by his name during his lifetime. Price came into possession of Bayes's papers after Bayes died; he worked up Bayes's theorem from the papers and sent it to the Royal Society.

[3]Robert Bailey should not be confused with his father, Arthur L. Bailey, who was one of the earlier proponents of credibility theory.

1959). Their study dealt with automobiles in Canada, but their conclusions with respect to credibility and experience-rating have implications for the health insurance area, and may be summarized as follows: (1) experience of one car for one year is significant, (2) the broader the rating classification system, the higher the credibility of an observation, and (3) the second year's experience increases the first year's credibility significantly.

As indicated by the first assertion, depending on the facts and circumstances, the limited experience of individual risks (or individual units of risk) can be material. On the other hand, as postulated by the second assertion, if there is a very sophisticated classification system, the individual risks have already been identified, and their experience becomes less and less important. Finally, assuming convergence, time series analysis may quickly classify a given risk.

What's new and different? What did we learn since 1918? Well, when Whitney was confronted with the problem of assigning credibility factors, he observed that it was "very, very subjective." While you still hear that today in some quarters, the new techniques are those which have been developed to resolve this issue.

A good deal of the credit for the development of these techniques is attributable to Hans Bühlmann. In his paper, "Experience-Rating and Credibility" (*ASTIN Bulletin*, 1967), he used least squares to evaluate the function

$$\min\ E\{(X_{n+1} - a_0 - \sum_{i=1}^{i=n} a_i\ X_i)^2\},$$

where X_i, the claim experience at time i, $1 \leq i \leq n$, is independent and identically distributed, with a mean of μ and a variance of σ^2.

His well-known conclusion was that X_{n+1} would be of the form $X_{n+1} = Z X_n + (1-Z)\,\mu$ where the credibility factor, Z, is $Z = n/(n+K)$ and $K = \dfrac{E(\textit{process variance})}{Var(\textit{hypothetical mean})}$.

Anyone who has studied for the current SOA examination on this subject is familiar with S. W. Philbrick, "An Examination of Credibility Concepts" (*PCAS* 1981, pp. 195-219), and SOA Study Note 422-34-89. Nonetheless, it is useful to revisit what he says and his conceptualizations. He summarizes by observing that the credibility of the current observation increases with: (1) increasing number of observations, (2) decreasing process variance, and (3) increasing variance of the hypothetical mean.

The first assertion is not surprising; the more relevant information, the better. To show the second and third assertion, Philbrick conceptualizes four individuals, A, B, C and D, shooting at a target, as shown in Chart 2. Their overall mean is located at "E." One of the four takes the last shot, as shown by the "X" in the figure, and the question is, who should the X be attributed to?

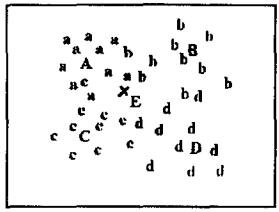

Chart 2

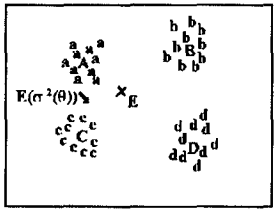

Chart 3

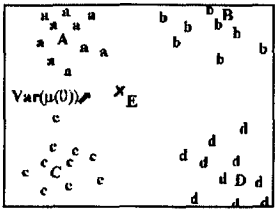

Chart 4

While a case can be made for attributing the X to individual A, it is difficult to make a highly credible judgment based on Chart 2. The problem is simplified, however, if the distributions are as shown in Chart 3. This follows since the process variance associated with A, B, C and D has been decreased ($E[\sigma^2(\theta)]$ decreases, where θ is the risk parameter for the individual risk). This causes their distributions to be in a tighter cluster and reduces their tendency to overlap, and, thus, makes it easier to classify the current observation. A similar result can be obtained, as shown in Chart 4, by increasing the variance of the overall mean [$var\{\mu(\theta)\}$ increases], which tends to segregate the distributions.

Lest it appear that credibility theory was investigated primarily by casualty actuaries, it should be pointed out that G. F. Hardy, in a correspondence to the *Insurance Record* in 1889, posed the following question. Suppose, in the simple case, that we are confronted with the likelihood of the Binomial distribution $p^n (1-p)^{N-n}$, where the p's are not degenerate, in the sense that they, themselves, have a distribution. What is the role of credibility theory when updating the distribution of the n's?

To put this issue in a more obvious actuarial context, we can write $(p_x)^n (1-p_x)^{N-n}$, where $n = l_x$ and $N = l_{x-1}$, and the tabular value $q_x = 1-p_x$ comes from a distribution. In this form, it is clear that the l_x's are binomially distributed, from which it follows that a convenient choice for the prior distribution of p_x is the beta distribution. See "A Bayesian Approach to Persistency in the Projection of Retirement Costs" [Shapiro, *TSA* XXXI (1979): 337].

Numerous other well-known life actuaries have considered the issue. For example, at about the same time as Whitney's study, Whittaker (of Whittaker-Henderson fame) wrote the paper "On Some Disputed Questions of Probability" (*TFA*, 1920) in which he addressed the relevance of Bayesian analysis to the actuarial area. As far as I could tell from the literature, he was the first one to use the target analogy to try to focus conceptually on the experience-rating and credibility issue.

The potential relevance of a Bayesian approach to credibility is appealing, since it combines past experience, current information, and judgment. A number of questions, however have to be resolved during the implementing of such an approach.

An important question has to do with the relative emphasis to be placed on the prior and the likelihood. That is, how should the weight be distributed between past and current data? This is the issue addressed by Whitney and Bühlmann and forms the basis of much of Chuck's presentation. A related consideration is the relevance of

published rates. Here we must be concerned with hierarchy problems and a well-defined model. Finally, what is the role of outliers?

MR. FUHRER: I'm going to talk a little more about the practical application of some of these things. The first question that comes up: Why do we experience-rate? In particular, in group health insurance, why do we bother? Now, when I talk about experience-rating, I'm talking about the process of using a group's prior claim history in determining the rate that the insurance company is going to charge that group in the future. Remember, the term "experience-rating" is sometimes used differently. Sometimes it's used to refer to setting the rate after the coverage has already been in force for a year, but that's not to what I was referring. Of course, the principle answer to the question, the main reason that we experience-rate, is because of competition. If we don't charge a group the most accurate rate as possible, then we are going to get hurt in the competitive marketplace. Those groups that we underrate will buy the insurance from us, and then we will lose money. Those groups that we overcharge will probably buy their insurance from one of our competitors. As you know, the group health market is extremely competitive. Most groups will get bids from competing carriers almost every year, and any significant price reduction will cause them to move. So, if you're not as accurate as possible in your pricing, you're going to be losing money, and if you're more accurate than your competition, you're more likely to make money.

Now, the second answer would be that there's obviously more equity in charging groups the more appropriate rate. I don't know if everybody cares about equity. In this age of regulation maybe it is important. Of course, some regulators have expressed the opinion that maybe experience-rating is not very equitable. Certainly in automobile insurance, it's well established that there's a need for experience-rating. Nevertheless, if you've had many accidents and you get rated much higher, you won't really like experience-rating, and you'll probably complain to your regulators that you're being overcharged and it's not fair. This is the type of argument that is going on in small group health insurance, where they've actually been limiting the amount of experience-rating companies can apply. I don't know much about this, because I work mostly with the larger groups. The concept behind the rules is that if somebody's health gets worse while we're insuring them, then it is inequitable to charge that person or group more money in the following year. Somehow the insurance company is on the risk for services in the future arising out of health conditions that started during the current coverage period. As far as I can tell, the insuring clause, at least in group health, does not put us on the risk for that. We are only insuring claims when there are services provided during the coverage period. I think that until we change our contracts to define our claims, so that if somebody gets sick while we're insuring the group, we'll pay for all the claims from that sickness forever, then we should experience-rate. This is the argument as to why experience-rating can take into account the health status of people in the group as represented by their recent claim history.

I did develop a formula for credibility by size of group, and I'd like to briefly go through it. The paper was in the *Transactions* [XL (1988): 387] a couple years ago. It may not be easy to understand, particularly if you are new to the subject. I suggest that you read the paper, look at the *Record* (16, No. 1: 1) "Managing the Bottom Line -- Group Excluding Multiple Employer Trusts (METs) where I gave a short

seminar on it, or review an even shorter article that appeared in the May 1990 *Health Section News*. In any case, I thought I'd run through it quickly. I think that my formula is a relatively good way of setting credibility levels. Here is the formula (the most simple version):

$$Z = \frac{k_1 + (m-1)k_2}{1 + (m-1)k_2}$$

Z is the credibility, there are two constant parameters which I've called k_1 and k_2, and m is the number of people in the group. Now you can use the number of people, but, as you know, most of us don't know exactly how many dependents are covered in a group. So when I did this, I used an m equal to the number of employees plus the number of dependent units. I thought of a dependent unit as sort of one large person. You could just use the number of employees for m, as long as you're consistent throughout and do your estimation of the k's consistently.

Notice that when m equals 1, the second terms in the numerator and the denominator drop out, and what we are left with is k_1 divided by 1 or k_1. So, in a sense, k_1 is the individual credibility. This is the major difference between this and prior formulas. In my formula, there is both an individual credibility and a group credibility: k_1 and k_2, respectively. As m gets larger, Z will increase and approach 1. Now let's look at what the k's mean:

$$k_1 = \frac{Cov(X_{i1}, X_{i2})}{Var(X_{i1})}$$

and

$$k_2 = \frac{Cov(X_{it}, X_{jt})}{Var(X_{it})} .$$

The X's are claim cost for a year divided by manual premiums or manual premium loss ratios (MLRs), not just claim costs. This appropriately adjusts between different people in the group who have higher and lower expected claims. By manual premiums, I actually mean the manual premiums by individual. Now, we don't usually calculate individual manual premiums in group insurance, but it's not very hard to figure out what they might be. We all have age factors, usually, by five-year brackets. So, you could just calculate the group's manual claims rate and multiply by the individual's age-sex factor. The i (the first subscript) refers to the individual within a group, and the second subscript refers to the year of claims. The k_1 is the individual credibility equal to the covariance between years one and two of individual MLRs divided by the individual MLR variance. We can estimate this covariance by using data on all the individuals in the whole group portfolio. This estimator uses the observed sample covariance and variance. It may not be technically the best estimator, but it's about the easiest one to understand, and it has some good properties if you have a large enough sample. Since we can use all of the individuals in all of our groups for this estimation, it is easy to have enough data. If you have 40,000 or 50,000 individuals covered, I think you'll get good answers.

The second factor, k_2, is defined in terms of one year's (which I've labeled t) claims. This can be any year for the estimation. It's equal to the MLR covariance between different individuals (labeled i and j) in the same group, divided by the variance. The process of estimating k_2 is straightforward. Remember that the covariance between two variables is just equal to the mean of the products of the variables less the product of the means. First we need all of the products of different individuals' MLRs within each group. We don't want MLR products of people in different groups. We're trying to see how much correlation or covariance there is between people in the same group. If it turns out that people in the same group have a lot of correlation in their claim experience, then the covariance will be higher. If it turns out that there's none, then it'll be zero. If it turns out that people have a tendency to have relatively different sizes of MLRs, even though they work for the same employer (go to the same hospitals), then k_2 would be negative. The process of calculation, summing up all of the products, might sound time-consuming (although for computers it's not too bad), but actually there's a little trick that can help. The square of the total of the MLRs for a group is equal (by the distributive law) to the sum of the squares of the individual's MLRs plus twice the sum of the product of MLRs of different people.

$$\left[\sum_{i=1}^{m} x_i\right]^2 = \sum_{i=1}^{m} x_i^2 + 2\sum_{i=1}^{m}\sum_{j\neq i}^{m} x_i x_j$$

where x_i are the actual claims for individual i in a particular year. If you then just subtract the squares of the individuals (from the square of the group's total), you'll have all the sums (actually twice) of the product of different individual's claims. Then we need to sum over all of the groups, divide by the total number of these products, and then subtract the square of the means. Now let x_{ik} be the actual claims for individual i in group k and m_k be the number of individuals in group k. Then we have:

$$Est\ \left(Cov(X_{it}, X_{jt})\right) = \frac{\sum_k \left\{\left[\sum_{i=1}^{m_k} x_{ik}\right]^2 - \sum_{i=1}^{m_k} x_i^2\right\}}{\sum_k (m_k^2 - m_k)} - \left[\frac{\sum_k \sum_{i=1}^{m_k} x_{ik}}{\sum_k m_k}\right]^2 .$$

This particular estimate of k_2 may not be very good. It isn't very robust and has a relatively high standard error. You may not get usable answers even with reasonably large portfolios. When I wrote the paper, I was extremely lucky in that the data that I used from Blue Cross/Blue Shield of Illinois gave me some very reasonable looking answers. Of course, nothing stops you from ignoring the data and just using, as Arnold said, a subjective number. At least try the formula first.

There was an alternative, older way of doing this. My formula will always give better answers. Some people may be still using the older method. If you look at the Society syllabus for examination G-422 (Group Health Pricing) there's a paper by Margolin in which he describes this other calculation of credibility. Basically, he uses k_1 formula, except that he uses the group claims and ignores claims at an individual

level. He takes all the groups that are approximately one size and calculates their credibility. For example, he would use all of the groups from 75-125 lives to calculate credibility at 100 and groups from 125-175 to calculate credibility at 150. I think he had a very large amount of data. My formula allows you to combine all of your data together to come up with your answers by size of group, whereas in his method, you can only combine the groups of a particular size to get each answer. You have a much smaller amount of data for each size, and you're much more likely to calculate ridiculous answers. We would normally expect the credibility to go up by size of group. That may not occur with a smaller sample in each size calculation. So, maybe you could do some smoothing. The Margolin method can give strange results one group in your portfolio had a very high experience in one year and a very low experience in the next year. This kind of negative correlation could throw everything off unless you had thousands of groups in a data set. If you're looking at larger groups you would never have this much data. My method gives answers using all the data, even from the smallest groups. In fact, you might not want to use the larger groups in your sample. Even my method can still give unusable results.

Let's look at typical values of the parameters so we can tell what's reasonable. In my study, k_1 was 24%. This seems pretty typical. I've spoken to three or four other actuaries who have done this on their data, and they've also gotten numbers of approximately that order of magnitude.

This has created a controversy. The value of k_1 is the credibility of a 1-life group. I've been at a few meetings or have overheard conversations about this in which actuaries thought that my credibilities were much higher than they should be. I haven't actually been able to confront somebody and find out on what this opinion is based. So, I'm not sure that I could actually characterize what the controversy's about, but I will admit that the credibilities were a little higher than even I thought they'd be. There was a general feeling, when I asked about credibility at Society meetings five or 10 years ago, before I did the paper, that groups of under 100 lives have zero credibility. In fact, a similar thought was expressed at one of the panel discussions at the Dallas meeting (*Record*, Vol. 16, No. 1) where somebody said that they'd found that there was zero credibility for groups under a 1,000 lives. I was really astonished by this, but he must have been talking about a different subject. I do have some ideas as to why actuaries thought that there was no credibility under 100 lives. If you go back to the original formulas for credibility and look at k_1, it looks much like a regression coefficient. In fact, it is a regression coefficient. The regression coefficient is the same as the correlation coefficient, except that the denominator is the first period's variance instead of the geometric mean of the variances from each period. Many actuaries are very familiar with material on linear regression such as that on Part 120 of the Society exams. In this material, it's said that correlations of under 70%, or 80%, or certainly under 50%, are useless and can be ignored. This may be true if you're trying to discover if there was a linear relationship between two variables that was obscured by random errors. In this case, if you have correlations down in the 15% or 20%, or even 40%, range, there would not be much evidence of the linear relationship. In credibility, if the regression coefficient is low (even 1%), there's no reason not to use the experience to modify the manual claim cost.

I have an intuitive explanation as to why the credibility would start at 24% and then go up as the group got bigger. We're using the claims experience as a way of measuring the health of the people in the group. People who had high claims one year are much more likely than the average person to have higher claims the next year. Thus, seeing an individual with a credibility of 24% is not a surprise. For example, if you were going to insure an individual who had major surgery last year and who spent five times the normal claim cost, wouldn't you want to charge that person maybe double the premium next year? Well, that's 25% credibility. We can look at individual health insurance. Nobody would dare rate individual health insurance with just manual rates. In fact, not only do they look at last year's claim experience on individuals, the individual health underwriters will actually find out all they can about the individual's health. If the individual health underwriter only knew what the amount of claims were last year and nothing else, they certainly wouldn't ignore that information.

Another reason why there may have been the controversy about higher credibility level is not only had the actuaries been saying that credibility was very much lower than mine, but they, of course, had told all the underwriters who had told all the salesmen that this was an official actuarial opinion. In the meantime, most of the sales and marketing people, I think, generally preferred rating groups with as much of the group's experience as possible. It's very easy to talk to a consultant, a broker, or the group about experience-rating. It's easier to explain than manual rating. The manual rating would appear to be a black box that has nothing to do with the employer, whereas the experience-rating looks very simple and straightforward. Experience-rating takes the group's claims, puts on the amount that hasn't been paid yet, the claim reserve, adds trend, and thus gets next year's estimated claims. But the underwriters said, "No, our actuaries tell us that your experience is not very credible. Instead, we're going to average it with our manual rates." So the salesmen are always pushing for lower credibility. Now I come along and say that, well, maybe the salesmen were correct. Many people don't really like to have to go back and admit they were wrong. Anyway, I do think, though, that it's important to understand the marketplace. One thing that we can learn from all of this is that the collective opinion of the market is often correct.

I estimated k_2 in the paper at about 2%. Of course, I haven't been able to estimate k_2 with other data, because of the problems I mentioned before. My feelings are that k_2 probably ought to be somewhere between 0.5-3%. I would really love to get data of this type from many companies and put it together so that we all can get a better idea as to what k_2 should be. I haven't succeeded in getting this type of data collected by the Society yet, although I have tried.

I just want to mention a few more topics, not in any great detail, and then answer some questions. The first one is the problem of determining pooling levels in group health insurance. This is probably a subject that is more important than the credibility levels for the larger groups. The formula for credibility never does get to 100%, but it does get very close. If you look at my paper, there is a method of calculating an optimum pooling level for prospective experience-rating. For this, you just need the same types of data items that we just looked at, but this time, calculate by using different limits on the amount of claims that you'll take from each individual. Then go through the regular calculation of those covariances, and pick the limit (pooling level)

that gives you the least-squared error. Now, once again, when I'm talking about pooling levels here, I'm talking about the ones that are optimum for rating purposes. The term *pooling* is used also for the refund or the retrospective experience-rating process. The levels I came up with are in the last table in the paper. Remember those are 1985 dollars, so you would probably have to multiply them by 5-8 years of trend. In any case, I would also be very interested to see what other people in the companies got for pooling levels if they did this.

Let me just go on to one more topic: stop-loss. First of all, I'm referring to specific (individual) stop-loss, although obviously you could use the credibility method to calculate the expected claims under an aggregate contract, but that's really no different than calculating it under the traditional insured case. Of course, you could treat stop-loss claims just as we did the other claims. For that purpose, we'll get very low credibility for smaller groups, so it occurred to me that there must be a better way of doing things. If you're familiar with specific stop-loss in various companies, you'll probably know that there are two very distinct methods being used right now to calculate specific stop-loss rates. One is where the company has a set of manual rates for a stop-loss. They tend to vary by, of course, the stop-loss attachment point, and by age and sex of the employees. We could call this the manual rating method. A second method that's used quite a bit is the insurance company determining what percentage of expected claims are incurred for amounts over various attachment points. For example, it may be 7% for $50,000, and it's probably around 2% for $100,000. Then the company will multiply the expected claims for the group by those percentages, and that'll then be the expected claims in the premium. These are two distinct methods of calculating rates, and I have been asked, which one should we use? I know there are many proponents of both, actually, and it occurred to me, when I was working on this paper, that it was possible that the answer was somewhere in between. We can use a credibility average between the two. The formulas are in the paper.

MR. BRENT LEE GREENWOOD: I view credibility as being somewhat of a premium-smoothing type of a process so that the rate variation from one year to the next can be somewhat minimized and really is somewhat dependent on management's objectives. Is management comfortable with a potential 30% swing from one year to the next, or does it want to minimize it and maybe have only a 10% swing in rates from one year to the next due to experience? I'm a little confused when you have 75% credibility for a group size of 100. The question is, did you do any analysis to determine what potential swing there is in premiums from one year to the next when you're using such credibility? If you do have a 30% swing, because you gave it such a high credibility, you may not be able to sell the group, but then it may go down the next year. So, is there any analysis on the impact on the rates and the variation from one year to the next, and what is that maximum variation that your factors come up with?

MR. FUHRER: I didn't do a study of the type you recommend. First of all, I disagree that credibility and experience-rating is mostly for smoothing of the premiums. The main purpose of experience-rating is to get the most accurate rate forecast as possible. Earlier, I explained why getting more accurate rates is important to insurers. Now as far as smoothing goes, it seems to me that the marketplace won't allow us to actually do more smoothing of groups' rates than was justified by the data. That

is, if we smoothed in the sense that we didn't relatively increase premiums as much as the claims data suggest, then we would lose money. Conversely, if we don't reduce the rates as much as the data suggests, then we will lose the group. This last result assumes there is another insurer bidding on the group that has the claims data and is using as high a credibility as ours. Of course, if we know that no other insurer has the claim data, we might as well charge those groups (with good experience) more. This strategy is dealt with in my paper. I'm also assuming that it is not the percentage of a rate increase that causes a group to leave, but rather the fact that they can find a lower rate.

If you were looking at some sort of captive, noninsured situation, you might start really being interested in smoothing cost from year to year, and then you might look at what the probable effects were in terms of smoothing.

MR. CHRISTIAN A. ULMER: First, I want to just say that I'm watching with interest Chuck Fuhrer's one-man crusade to change credibility theory, and I appreciate it. I wish you well. My question is probably at least a little related to the last one. The great achievement of *Actuarial Mathematics* over the old Jordan *Life Contingencies* textbook was that not only did we look at expected values, but we have variances and confidence intervals. Credibility formulas basically give you an expected value for this group as opposed to all groups. The question would be, how would we arrive at a variance calculation for a given group? Then we could load it so that we had 80% probability of premium exceeding claims.

By the way, we may not choose to go ahead and rate the case with the correct credibility or the variance because of smoothness considerations or because the regulator just told you that 50-life groups can't be rated that way. I think it would be useful and it would further knowledge if we did have such variance values.

MR. FUHRER: Thanks for the support. Actually, I am not trying to change credibility theory. In fact, my formulas are a direct result of all of the current thinking in credibility theory. I am merely on a crusade to get the practicing group insurance actuary to use these results.

The variance question is a good one. First of all, with the linear least-squares-type credibility factors, it's a relatively straightforward calculation to calculate the so-called residual variance. I didn't mention it in the paper, and I think that would be a very good thing to do. If you do pull out the exam 120 text, you probably could figure out how to go about doing that. If not, I'd be happy to talk to you about it sometime, but it makes me wonder if I shouldn't actually try and set that out some place. I think it's an excellent thing to do.

FROM THE FLOOR: Could you say a bit more on this subject?

MR. FUHRER: Yes, there are some other ways to go with this. You might think that you should look at the group's own record of how much it's fluctuated so that you could actually estimate the variance based on the group's own variability. In other words, develop a credibility-type approach to actually estimate the group's variance as opposed to just looking at the residual variance, and, in fact, there has been some work done on that. Hans Bühlmann, in his book *Mathematical Method in Risk*

Theory, actually developed a linear credibility factor for the variance of the individual risk by using the sum of squares of the prior periods' experience. I did use this at one of the companies I was with. I was looking at aggregate stop-loss premiums where we really need badly to know how much fluctuation there was in a particular group. I actually had the underwriters calculate this variance credibility and use the resulting variance to adjust the aggregate stop-loss premiums. So, there is a good application of this.

If you want the whole distribution, not just the variance, some work has been done by treating the distribution as a Bayesian posterior distribution. You could check: "The Credible Distribution" by W.S. Jewell in the *ASTIN Bulletin* 7 (1974): 237-269.

MR. CARL D. SMITH: Would you expect that as the health care trend takes its toll over time. I'm thinking here in particular of the utilization component of trend, that this would cause the numbers to change, going through your calculations?

MR. FUHRER: Well, that's really interesting. I guess I hadn't thought about it at all.

MR. SMITH: I guess it just seems to me that as claims become more frequent, more credibility could be attached both to the individual and to the group, but I'm speculating.

MR. FUHRER: I really hadn't thought about that at all. Your suggestion sounds good and has some intuitive appeal, but I'm not sure.

MR. SHAPIRO: That is one of the implications of what Bailey and Simon observed, and it is the essence of A. L. Mayerson's paper, "A Bayesian View of Credibility," *PCAS* (1964), p. 85. Of course, convergence is an important issue. If a parameter of a valuation is unstable, it often is unclear whether a high credibility can be attached to an updated estimate. In some instances, such as in the case of certain decrement rates, economic and other exogenous factors may have limited effect, and convergence can be assumed. In other instances, however, such as with factors that are dependent on the economy or regulation, convergence may not occur. The point is that your hypothesis is correct, other things being equal.

MR. KENNETH K. LAU: How do you take the turnover rate of the company into consideration when you have a lot of turnover in your group? How will it affect your credibility? In my work, frequently I'm not so much concerned about a company's own credibility, but the question usually is whether the allocation of expenses into smaller subdivisions of the company are accurate. Sometimes with a spin-off situation that becomes an important question.

MR. FUHRER: First of all, I might be accused of having you as a shill. In the paper, in the *Transactions*, there was an adjustment formula to the formula for turnover. Here is the new formula: with p the persistency. The assumption in the first formula

$$Z = \frac{pk_1 + (m-p)k_2}{1 + (m-1)k_2}$$

is that there's no turnover, which is clearly not true. Of course, you could estimate

the parameters by using all of the individuals, even those who leave the group. This would give you a lower credibility than using only those individuals who stay insured. If you use all of the individuals, then your credibility will be correct if the group that you are looking at experiences about the same level of credibility as was in the data.

MR. SHAPIRO: One of the issues that has not been addressed is the role of what-if analysis. In this regard, it is worth mentioning that credibility models lend themselves to simulation. That is, you are given prior distributions and sampling information, and, from these, you can construct posterior distributions. A stochastic simulation of this process can be used to develop confidence intervals for the cost estimates.