Session 101TS
Risk Theory Applications To Experience Rating Group Health Insurance

Track: Health
Key words: Group Health

Instructor: CHARLES S. FUHRER
Recorder: CHARLES S. FUHRER

Summary: The practicing actuary needs to set credibility levels to be used by the underwriters to set group health insurance expected benefits. Instructor explains how to set those levels and explore various real world problems that arise. Both small group (1–200 lives) and stop-loss insurance will be covered.

Mr. Charles S. Fuhrer: If you attended Session 9 you will have been brought out from the before-credibility era, or, as people like to refer to people who aren't up-to-date, as being in the Stone Age. However, the effect as far as group health insurance in my opinion was to only take you up to the era of the Roman Empire. Consequently, I've used Roman Numerals which, as you will see, are incredibly clumsy, as well as the fact that if you just use the credibility theory as presented in Session 9 ("Believe It Or Not"), you will find it to be somewhat clumsy in your everyday work.

Let me do a little housekeeping. First of all, I gave out no handouts. That was quite deliberate. I've found that generally when there are handouts people look at the handouts and take them and go, and I would prefer that you pay attention. In fact, if you would like to get a copy of the slides, I would be happy to send it to you. I am not in the Stone Age. Now, it'll also cost you. There's a price involved in getting these handouts which probably is an overestimate on my part of what they're worth, but, nevertheless, there is a price. The price is that I assume that if you want the handouts, you're going to go to your company, and you're going to do some work and get a hold of some data and do some of the sums and get some of the results for the parameters on the formula that I'm going to be working on. I am going to ask you to send me a list of those sums so that we can pull them together

---

and maybe get some answers as an intercompany study so that maybe we can get some answers that may not be available in all but the very largest of companies.

I dropped the word credibility from this session. I'm not sure why I did that. I'm getting a little sick of talking about credibility. I've been doing it for about 10 years, ever since I wrote this paper in *TSA* XL which was in 1988. My discussion will follow somewhat the talk I gave in MCXC which was in Dallas and is printed in the *Record*, Volume XVI.

Let's talk about least squares credibility. First of all, in Session 9 they talked about a loss function which is used to determine how to make sure your formula's optimum. The usual one that is used is least squares, and I guess there's been a general feeling in the statistical community that least squares was something that was done because it was easy to calculate. We don't need it anymore. We should use something like least absolute deviation, but least squares has a lot of appeal in its simplicity, and, in fact, it is sort of what people normally think of, despite the square. For example, the mean is the least squares estimator of population. The median would be the least absolute deviants which, of course, is used quite a bit, too. Then, of course, try and balance a pen at the median, and you'll see it won't balance, whereas the least squares one will.

I guess we could try some other loss functions, but I think I'd like to stick with the ones that produce means, at least until we all learn those, and then maybe we can worry about some other ones. Now, the next thing that we use is a linear formula. This can be justified in statistics by coming up with some sort of special distributions and assuming it matches those. But basically I like to use the linear formula because it's simple, easy to explain, and we can still make it a good estimator. So, we'll stick to that also. In Session 9 they gave you the formula N over N plus K, and my response to this is no way.

Let me just explain in words why I don't like it. First of all, the N in that formula originally meant years or insurance periods, if they weren't using years. The use of it for the group size was never really established, despite what you heard in Session 9. In fact, to say that it could be used for the group size basically involves the assumption that each successive member of the group is almost to be treated as exactly the same risk, just like you had more years of data. Now, why not consider a group as a bunch of people with the same risk? Well, let's take a look at this. Suppose you had a one-life group. How much credibility would you give to the claim experience? Now, giving zero credibility would be equivalent to not underwriting that person. It would be saying we're going to ignore the past. We're just going to charge them the manual rate.

We all know very well that we want to underwrite individuals when they apply for insurance. The individual department would lose a lot of money if they did not. If we were faced with the problem that the only information we had was the claims data, then we wouldn't be quite as happy as looking at the medical conditions. However we would still vary the rate considerably, wouldn't we? For example, if the person had double the normal claims amount, you can expect that person probably had quite a bit of medical care in the last year. That person is much less healthy than an individual who had not had any significant amount of health care, and, therefore, we would want to charge them more. Similarly, a person that had not seen any health-care providers, or at least had not met the deductibles, is probably more healthy than the average person. I have found repeatedly in every single data set I've ever looked at that the credibility of one person runs in the 20-30% range. That means a healthy person, doesn't have any claims last year, assume is moderately healthy, would get a 20–30% discount, and a person with double the claims would get a 20 or 30% load.

Now, if you're going to use N over N plus K, and when N equals 1, it's going to be at least 20% credibility. That means that when we get to a 100-life group we have at least 96% credibility. In my experience 100-life groups bounce around a lot from year to year. If we just use their last year's experience and ignoring manual or we weighted at 96% and 4%, then I think it would be the wrong answer. Well, what's the problem here? Well, for a long time, back in the ancient 1980s, I looked at this problem and tried to fit it. I noted that if you just use the 20% and thought of a group as just a bunch of individuals, then, sure enough, the credibility wouldn't get higher as the group got bigger. We all know that a 500 or 1000-life group starts to have a significantly more than 50% credibility. Maybe a 10,000-life group. I don't know. But certainly the credibility goes up. We start believing the experience, and it becomes quite predictable. What's going on here?

The problem is with the model, that there's both an individual correlation in successive years between persons, and there's also one in terms of the group membership. There was a paper published in the 1980s by Bill Jewel. I first became aware of this because of a paper written by Gary Ventor, and they suggested that if you're trying to estimate next year's mean for an individual that you basically needed a two-dimensional credibility. You needed one credibility, Z, applied to that individual's claims and another one applied to the average of the group, and, finally, any residual non-credible piece you would use the mean or the manual rate for. This particular formula worked just fine, as far as I know, but we're not interested in predicting the claims on a particular individual. We're only interested in the claims on the whole group. What we needed to do was add up each of these for all the people. The result was that this term here, which is the

individual's claims this is (Z_1 x X_1) + (Z of the group times X of the group) which is the group average claims.

Mr. Fuhrer: If you take this formula and add it up for all the group members, then the sum of these just becomes the group average. So, we end up with a single credibility factor, but the factor itself is calculated differently because we have to figure out how to calculate these two Zs and then how they add up together. Now, when I published that paper in 1988 I'd never seen such a formula, and I have not seen it appear in print any place else. Now, just a little aside here. For the purpose of this we're going to use the method-of-moments estimator which means basically we take the sample covariance and substitute that in for the covariance of the population that we're trying to estimate.

In the mid-1900s there was a calculation formula used where we take the mean of the products and then we subtract the product of the mean. Just a little algebra would convince you that's the same, and since I'm going to talk you through this calculation, I thought I would use this somewhat simplified calculation which, of course, isn't really necessary if we have computers. Here's the big formula. The group credibility, instead of (N over (N plus K), is K_1 plus (M minus 1) times K_2) divided by (1 plus (M minus 1) times K_3). Now, I used M here to mean the number of members in the group to distinguish it from N which is the number of years, in my opinion.

Let's look at this formula for a second. If M equals 1, then this reduces down to K_1, and the K_1 is that individual credibility that I spoke of. As M gets larger, the formula becomes more and more like the ratio of K_2 to K_3. Therefore, I reasoned that it's likely that the parameters K_2 and K_3 could be replaced with a single parameter. Now, when I did this work originally I was working at Blue Cross/Blue Shield of Illinois, and I extracted some data from about 1984-86. I got some really good answers. They looked fine. It gave me a K_1 of about 25% and a K_2 of just under 2%, I believe. Those answers looked real good, and, in fact, K_2 and K_3 were almost exactly the same. So, that felt good, too.

Looking back at the size of my data set, it wasn't very big. I've looked at some other data sets at my subsequent employer's, and the data and the results did not come out quite as neat. I always got good answers for K_1, but the method-of-moments estimator for K_2 and K_3 can give quite a bit of variance because they are at the group level. You literally need a thousand groups to be very comfortable with it, and there are not too many carriers that have over a thousand group. It looked to me like I was just lucky, and, in fact, the 2% looks a little high. If you put 2% in for K_2 and K_3, you'll see that you're getting up there in the 90s pretty quickly, and that didn't hit my preconceived notion that the credibility should be quite that high so

quickly. But part of the problem here is how we define M or members, and that I think is a Blue Cross term. Individuals covered might be more of an insurance industry term. But even when I did this work originally, it wasn't actually individuals covered, and I'll get to that in a minute.

Now, how do we go about calculating this? Let's determine the parameters in the formula. Now, this is a nine-step method. The first thing we need to do is obtain some claim data. We need at least two years, and then it should be probably two years of incurred claims. We could probably get a good estimate of incurred claims about three to six months after the second year, but we still might want to put some sort of an ultimate claim reserve on it. At least in theory we could maybe have used two years of paid claims hoping that more or less any ups or downs in the effect of using paid would cancel out. Another possibility I'll get to later, and that is to actually use the unreserved claims from year one which, of course, seems counterintuitive because those claims we already have, but I'll explain at the end why I would think that might be worth looking at.

The next thing we have to do is get the claims on a member-by-member basis for at least two years. A lot of us can't get that. When I did this work at Blue Cross/Blue Shield I was able to separate claims between the subscriber or the employee and the dependents, but I was unable to isolate as to which dependent. I treated a dependent unit as a single member and then went ahead and considered the employee as another member, and I used employees plus dependent units as my member count. So, that may have partially explained why the 2% was as high as it was, because presumably if I had used actual number members, then I would be multiplying by a bigger number, and, therefore, I should have gotten a smaller value for K_2 and K_3.

Another way to do it would be to merely combine all of the children and consider a children unit. I think the greatest accuracy would be is if we could identify each particular person. Then the next thing is we have to identify the group. This is not done when we look at various claim studies. People have a tendency to just get the data and get the exposure counts and then divide, but we actually needed it identified by member and by group. The next thing we need is some sort of exposure information. That's the membership, the number of members in each group. Surprisingly, this is not as critical as it sounds. We're going to be looking at the relationship between claims in successive years. We have a count of those people who have had claims just because they appear in the claims data set. So, the only thing we're missing is how many zero claims there are, and it turns out that's not real important in terms of our answer. Nevertheless, if it is available, I recommend using it.

Do we want to exclude members? For example, there are two ways of doing this. One is to only include members who were covered throughout the whole period. Then you can use my handy-dandy formula for adjusting for a turnover where new people come in and out of the group which appears in the *Transactions* article, or you could just assume that every group has roughly the average turnover that was in your data set, include all members, and then don't use the adjustment. The adjustment, by the way, is down. If new people are coming in and out of the group, that lowers the credibility. If you want to use all of your members and not exclude the ones that came in and out, you'll get lower credibility to start with. Presumably you won't be as accurate because you won't be using the actual turnover of the particular group you're looking at. Of course, that itself may not be very credible. I'm not sure which way to go on this. I put the formula in the paper, but then I did not exclude the people who were not there for two years when I handled the data in any of the companies I was with. So, you can go either way on that.

The next thing that you'd like to have is some information about the plan of benefits, the age and sex of all the people in the group, but it might be enough to just have the age and sex of the claimants. The next thing we do is we adjust the claims. The big adjustment here is we want to get rid of any factors that we already know about and we've already adjusted for in the manual rate. So, we determine a manual rate for each individual or each member. Now, that's not a typical thing that group actuaries do. We don't normally come up with a manual rate by member, but I would maintain that it's relatively easy to do that. We just take whatever age and sex factor we would apply to that particular member and then multiply that by whatever factor we would use for the other things that are on the group basis. We divide the claims by that manual rate to get them out of it so that we end up with a rate that is kind of adjusted relative to the manual. They're manual loss ratios. This should adjust for any relationships that are already there. When some statisticians have looked at this they preferred using an additive model instead of a multiplicative model. I don't really know which is better. This seems more natural and easy to work with.

The other question is whether we're going to pool large claims, and I use the word pool very loosely. What I just mean is exclude them from this data estimation process, not in terms of how we're going to rate the groups. I would recommend that you take out any really jumbo claims from this. They may have a tendency to push the data a little too far in the direction of too much individual credibility or too little, depending on whether that claimant had claims in successive years or not. It's easy in this type of analysis for one claim or a handful of claims to kind of dominant your answer. What's too much? I don't know, 100,000 or 200,000, something like that if it's your whole company data you're using.

Now we get to the actual calculation. The first thing we calculate is the average claims per member or average adjusted claims per member, and this is pretty easy. We just sum all the adjusted claims in each year and divide by the member count. I've designated these as upsilon sub 1 and upsilon sub 2. This is sometimes called the per member per year value.

Determine the variance of the members' adjusted claims. We sum the squares of the members' claims, and then we divide by the member count. Remember in this formula we have to subtract the means in order to get the right answer. The next step is determine the covariance. That's the covariance between each member's claims in successive years. We take the product of year one's adjusted claims and year two's adjusted claims for each member. We then sum these products over all of the data, all the groups. We divide by the member count, and then we do the mean correction of subtracting the product of the two means.

So far all we've done is normal calculation that one would do if you were determining the correlation or regressions of people upon people. The next step which gets to something new is to determine the group variances. This is not the variances of the group. This is actually something I made up. It's not even really variances. It's the covariance of different people in the same year who have something in common. They're in the same group, but they're different people. You can read the paper to determine what it is, but here's how to calculate it. First you calculate the total of each group's adjusted claims in both years. Then you take the square of the group totals, and we only need year one for this calculation, and you sum those squares. Then you subtract the member squares total. The reason we do that is we want the products of all the people's claims or adjusted claims in year one who are individuals but appear in the same group. Multiply the group totals together, and you'll get the all the products of all the people. Then you subtract the squares of the people within those groups, and you'll end up with the products of the different people. I don't know if that's clear, but you'll have to work it out.

For a member count we need to actually divide by the number of products. The member count here is the sum of the squares of the counts in each group minus the square of the total count. The sum of the counts squared in each group minus the square of the total count, and that's the denominator. Then we have to subtract the mean square from that average. The group covariances is basically the same thing. We take the-each year's total of the group's adjusted claims and multiply those together, sum those products, then we subtract the member products, divide by the special member count which is once again N squared—the sum of the squares of the group count minus the square—the total squared, and then we subtract the product of the means to do the regular mean correction.

The last step is to set K_1 equal to the member covariance divided by the variance. K_2 value equals the group covariance over the member variance. And K_3 equals what I call the group variance over the variance also. You should have three numbers that look like 25%, 2% or less, and 2% or less, with the last two being the same. This is not necessarily going to happen, and what I would like to see is people send me these sums-the total counts, the total counts squared, that's with the group, the group total counts, each one of those counts are squared and then added up. I would also like to see the other four sums, the sum of the claims in each year, so that's two more, the sum of the squared of the claims in each year, the sum of the group squares in each year, and the sum of the group products.

Then we can put these all together because we may not get very good answers for K_2 and K_3. The second time I did this I got negative answers which doesn't seem likely to be the correct thing to use here. If you look at the properties of what went into these things, you'll see that it's very dependent on what kind of groups you have in there, and it could easily jump around. I won't use the data for anything except I'll pull it all together from all the companies that send it to me, and then I will send it back to people that contributed as to what answers we got. I'm hoping that the effect of piling it all together, even though we're using different manuals, will be useful.

In addition, I'm working with the Large Claim Database Committee, and we're going to try and use the data that's submitted to that to kind of get some intercompany answers. Now, of course, we're asking them to collect all claims, not just the large ones, and the new large claim request for information will not only ask for large ones. Presumably that will be easier because the companies won't have to cap the claims at a particular value. They can just dump their whole data sets. We'll also have to have a group identification put on that. But that's really the only differences, and so that should be relatively easy to accomplish.

Mr. Roger T. Schacht: You talked about comparing the claims for a member from one year to another. How do you handle the situation in, say, a fast-growing Health Maintenance Organization (HMO) where you have a lot of members that were only in one year? Does that cause any problems?

Mr. Fuhrer: The question was what do you do if a lot of the members have not been in force for the full two-year period? My belief is that if you don't have a significant amount of data where they're in force for the full two years, you might as well not do this. You're not going to get very interesting answers at all. The whole point of the study is to determine how one year's claims helps you determine the

next one. If you don't have two years of data, you're never going to be able to answer that question.

Mr. Schacht: Is there any value to looking at this by component, inpatient versus physician versus pharmacy or anything like that?

Mr. Fuhrer: The question was is there any value in separate the claims out into types where you have, say, for example, hospital inpatient, hospital outpatient, physician, in-hospital and out of hospital, specialists? I mean we could have six categories or a hundred. I believe there's a lot of value in so doing. The disadvantage is that the size of the databases that we'd need to get good answers increases further, and I have not looked at it that way. I suspect that we don't really have good claim distributions when we're just doing pricing that break down the claims that way. At least I haven't seen any that look very good. I'm not sure how much would be gained by that, but I welcome an opportunity to do that if I look at some large data sets. I think on the large claim study they're planning to break it down into large categories. We may get an opportunity to look at that on that basis. Presumably hospital claims in one year, at least on an individual basis, tend to lead the higher claims in the following year than maybe just physician claims. There may really be something there worth looking at.

Ms. Kathryn L. Cole: I don't know if this question is going to be lucid, but what about the difference in experience between the first duration and later durations? You're trying to predict some kind of credibility, but you're using two years of claims to predict what? Just two plus or first year or second year? Do you split it between one and two plus?

Mr. Fuhrer: If I understood the question, I think it was what durations we're looking at and what the goal of this kind of credibility is. Are you also referring to durations in individual underwriting when you say duration?

Ms. Cole: Well, I'm primarily thinking of small group which is 2-50 employer level underwriting, but our actual-to-expected on first year is a whole lot different than two plus.

Mr. Fuhrer: Okay. So, you are referring to years.

Ms. Cole: Since issue, for the group, though, not for the individual.

Mr. Fuhrer: First of all, the attempt was originally on the group model where I was assuming group underwriting and the problem of what kind of renewal increase to get or what kind of rate to go out with when you get experience from the prior or

current carrier. That would tend to indicate that we're talking about over 50 or even over 15 companies are only doing individual selection above 15 or 10. I was thinking in terms of the group problem where there's only group underwriting involved. And the goal was to look at two years of data, see how the second year varied as a result of the first, so that companies, when they were doing renewal underwriting can and want to use the credibility of the first year's claims in adjusting the rates for the second year. With that said, again I don't know how well this would work if you're just sort of ignoring it. We might want to look at all first duration groups and see what kind of credibility there was on the second and future years.

Ms. Cole: I guess that's what I was thinking because if you just take two years of experience, you have all duration. For some of them, you're looking at the third and the fourth. With others, you're looking at the first and the second.

Mr. Fuhrer: The method I suggested just now would try and utilize all the data, put it all together, and then say that the durational factors I'm using are, I believe, and therefore I can do that. It certainly would be a better method to separate out the duration and to actually have one data set for first duration experience and another data set for first and second, etc. I just wonder how many companies would have enough groups to be able to come up with reasonably good answers for that.

Ms. Cole: You said something about the number of groups. Does this method work for small groups—we have over 4,000 small groups-or were you really just thinking this was for over 50 primarily?

Mr. Fuhrer: I suspect that this method works the best for the smaller groups. I have not actually had a data set where I had enough tiny groups to test that out.

From the Floor: We might be able to help you there.

Mr. Fuhrer: Right. Four thousand is quite a few. Recently my company got larger because there was a merger, and so we may have enough to be able to look at it also. Most of those groups are subject to group reform where we couldn't actually use it for rating, but we might be able to extract the data for use in rating the over 50.

Ms. Delaine B. Hare: I was looking to try to apply this to a disability coverage, and a couple of differences in disability that would make it difficult to apply this would be (1) you have a very low incidence rate coverage where probably, even on a short-term disability policy, 93 out of 100 of your members are not going to have a claim in any given year. The number of members where you have no claim in the

first year or no claim in the second year, is going to be very high. (2) A complicating factor would be that if you had a claim in the first year, then you're probably not exposed to the disability risk in the second year because you're still disabled. I was wondering if you could comment on how applicable this formula might be given those considerations.

Mr. Fuhrer: The question was how does this apply to disability income? Did you say long term or short term or both?

Ms. Hare: Probably both, maybe more so in short-term disability because you do have a little bit higher incidence rate there.

Mr. Fuhrer: The main reason for developing this formula as opposed to the ones that you saw in Session 9 and exist in the literature was the very fact that you have a lot of claimants who stay in the group. As soon as you start looking at a life insurance example or, in most cases, long-term disability, you don't need this formula because the regular N over N plus K works just fine. Now, short-term disability, it may come up, and I have not looked at that. It doesn't seem like people are too interested in short term, though. I'm not sure why. Perhaps because it's usually written with coverages that are more expensive, and it's relatively predictable. In any case it seems like it might be of some interest on the short term, particularly if it's very short so that typically the people would have short claims and then stay in the group and then are more likely to have short claims in the following year. But there's no reason to use it at all for long-term disability or life insurance where the N over N plus K works just fine. That's not to say there aren't some interesting problems there, too. For example, on long-term disability you almost never know what your incurred claims are, and so there needs to be an adjustment downward for the credibility of the claim experience because you're estimating it. I did a little work on that which I presented at the Washington meeting last year. However, I have not had time to write that up, but basically I did cover in there how one might adjust the credibility downward for long-term disability due to the long runout or the disabled life reserve.

I think the next part may stimulate even more questions. That is what are some of the practical considerations in using this? The first one is what happens if K_2 is not equal to K_3? In one data set I looked at, they were way off. K_3, I believe, and I haven't really tested this, the estimation I think is a little more stable and reliable. At least in one data set I got a pretty good answer for K_3 and just a terrible one for K_2. I just set K_2 equal to K_3, and that seemed to work. If you don't get good answers, you can always guess. A guess using this formula is more likely to give a reasonable optimum kind of credibility levels than using any other formula, in my

opinion. Finally, maybe we can get some intercompany data together and publish some sort of baseline values for this.

How are we going to use this? It's a simple formula. In my experience, though, if I try and give it to the underwriters, they scream. They don't want a formula. The formula requires two multiplications, two additions, and a division.

We could build it into the program that they're using. A lot of them don't trust programs either. Perhaps they're still in the B.C. period. A lot of times companies want tables, and, sure enough, they'll print out a table, 100-120 lives, 50%, 120-140, 55%, and this is certainly not crucial. The difference in the rate on a particular group is not going to be that much, and the amount of rounding that occurs in so doing is reasonable, and I see nothing wrong with that sort of tabular approach to things.

A bigger problem is to what extent you can give the underwriter some feel as to how they might adjust that based on a particular group situation, and that's going to be really hard because I'm not even sure what those considerations are. I think that if a group's data wasn't somehow tainted or we weren't sure about it that you ought to reduce the credibility. That makes sense to me. I think if you have to make an early call without getting the reserve in, then that ought to reduce it. Certainly if you have less than 12 months of incurred data, that reduces it. In the paper I did put a formula in there for what to do with less than 12 months. It's basically just an application of N over (N plus K), except that I applied the N to years or partial years into my formula. That wasn't very satisfying because I'd like to be able to say, how does eight months of data really relate? In order to do that we would actually have to take eight months of data for year one, etc. Another thing that was brought up in Session 9 was when you get to 100%. Clearly, if you're at 98%, you might as well call it 100%.

My formula, as well as the ones mentioned in Session 9, never actually get to 100%. If you're close enough for it to not make any significant dollar difference, then I would assume that it would be okay to just round it to 100. However, Session 9 brought up what I think is a more profound question, and that was what happens if there's competitive pressure to change these things, and what happens should you give in? How much is it costing you? The answer there is you could actually calculate, well, on the average I'm going to be off by this much, and so this is what it's going to do to the bottom line if we deviate from these credibilities that are optimum. Nevertheless, small deviations are not going to be very important because we're not even 100% sure if we have the parameters right. So, it's not going to make a big deal of difference. A lot of times the credibility that's used in the industry has gone a certain way because a lot of companies are doing it that

way, and the clients, the brokers, the consultants, have gotten used to certain levels and certain concepts, and to the extent that we're going to come up with something different, then we're sort of going against the stream, but that difficulty is not necessarily client driven.

In my experience what I've seen occur is that people had low credibility factors for under 50 lives. The underwriters were used to those. The salespeople, of course, wanted to give more credibility to a group because it's easier. For the renewal sale the manual rate is coming from some sort of black box in the company and is relatively hard to explain, and the clients and the consultants are quite capable of saying, well, that applies to your pool of business, but it's nothing to do with my group. They tended to push for a lot more credibility. I've overheard the conversation many times when it came to like a 50-life group—a salesperson said that it should be 30–50% credible. The underwriter said that their actuary told them there's no credibility at all at that level. I had to bite my tongue, which I don't do very well, and say, well, wait a minute. I think maybe 40% or 30% is quite reasonable for a 50-life group. When you do that to an underwriter, when they think that actuaries believe a certain thing, and then you come along and tell them, well, that's not true, and they've been telling salespeople a long time, they may not be your friend after that. That's going to make it even harder to work with.

I have run into this situation a number of times. I think the bottom line is that with any rating thing the closer you can come to what you think is optimum, the more likely your company is going to have a competitive advantage over companies that are not using it, therefore, the more likely they are to make money. It seems to me that, at least under certain circumstances, you need to insist. On the other hand, let's not quibble over a few bucks. I don't have a lot of advice, but it seems to me that's one of the advantages in putting in a table for them. At least you get a little easier buy-in for the other things you're trying to do which is the gross change in the amounts.

Another issue has come up, and that is how much this matters for HMO versus non-managed care. I don't know the answer to that. I guess I would be interested if companies could indicate on the data whether it came from indemnity or HMO to see if there's any difference. Now, of course, if you're not getting the actual encounter data, the actual claims per person, because your HMO is capitated, then you can't really do this. So, that kind of creates a problem.

Mr. Steven P. Clay: As I'm thinking about all this, one question I sometimes get is in relation to capitation from physicians. Why should I accept a capitation on *x* number of members? What kind of credibility does that have? I'm sitting here trying to think, would this be directly applicable? Could you comment on that?

Mr. Fuhrer: The question was how this affects negotiations with capitated providers?

Mr. Clay: If Blue Cross had a primary care physician group that had 800 members assigned to them, then we could use some sort of formula like this to say that there really is a lot of credibility in this group. You shouldn't be in this group of members. You shouldn't be too concerned about the variability of the risk.

Mr. Fuhrer: I'm not sure how to respond to that. I think that this particular method would work. It might give you an optimum answer in terms of the cost for those 800 insureds that were being capitated. Therefore, if a particular physician group could logically come in and say, we have worse experience on our 800 people than other physician groups not because we utilize worse but because we just happen to get sicker people. There certainly would be some credibility to that. On the other hand I'm not sure how to take it apart and know what has caused their experience to be worse. Possibly if we were to start looking at conditions only with some sort of objective severity put on it and looked at the credibility of that, then that would give us some answers. Certainly 800 lives, though, they could very well be sicker or not as sick.

Let me just point out that you're talking here about individual selection of providers. One of the reasons the credibility goes up with the size of the group is because there's assumed to be some sort of group selection going on. There's something inherent in the way the group either selects new employees, continues to retain employees, or something in the health conditions inherent in the location of that group, the type of occupations that the group has and the types of providers that are near the group in terms of where the people live and go to them. These are the group characteristics that we don't know about when we rate groups that we don't adjust for in our manual. I would suspect that when you had a group of people that all went to a particular provider who individually selected those providers you don't have the same types of things going on. On the other hand, you may have something even more remarkable in that only a particular level of health a person tends to select a particular provider group or on the average they do. Obviously people live near them. That's one consideration. This is an entirely new way of looking at it, and if we wanted to do that, we would have to re-estimate these Ks using those people who went to physician groups. I hadn't even thought of doing that. This is very interesting.

Mr. Van Allen Jones: Question on the methodology. Is there the possibility of some bias, given that you're selecting groups as each company has had insured for two full years, and, therefore, they've had to go through a renewal in the middle of

that, and their choice of staying or going would be related to what credibility formula renewal rating practice was in place?

Mr. Fuhrer: The question was whether by excluding groups that had not been in force for the full two-year period that we were introducing some sort of bias into the answers. I think the answer to that is no. I think that whatever operates to cause groups to leave during the period is not directly related to this question of how much members within a group are correlated with each other which is what we're really trying to measure with K_2 and K_3. With that said, though, I think that there are some other ways of doing this that might make you feel more comfortable. For example, we could look at all renewal years that, say, ended in 1997 as our first year, and all renewal years that ended in 1998 as our second year, and the advantage of that would be that we would at least know that we had a block of business where the years that they stayed with us were very coincidental with the decision to continue coverage in the middle of it. In other words the rate stayed the same during that period, etc.

I guess there still would be the objection that we're only selecting those groups that decided to stay with us at the middle renewal. Once again that would tend to probably kick out those groups that got high rate increases. I don't see a reason why groups that have high rate increases have any more correlation between their people. Maybe less. I don't know. It's a good question. Unfortunately, I don't know of any way of avoiding that particular problem. It seems to me a bigger problem here, though, is the self-selection of the individuals. Some of them are coming in and out of our plans. I really don't know how to adjust for that. I forgot to mention one other adjustment that comes to mind that might be more important than separating the claims between the hospital and the physician, and that is that this assumes that we don't know who actually stayed in the group and who actually left. That if there are any high utilizers who had a huge part of making a group high claims, and then they turned around and left the group, now that we're forced to rate this group we probably should not use their claims in determining the answer, and so we probably ought to figure out how to adjust for that.

One way would be to not use their claims in the portion of the credibility that had to do with the individual covariances but still use it in terms of the group characteristics because the fact that this person blossomed in this group is some indication that something's wrong or that people in the group tend to be less healthy. I didn't do that in the paper, but I have looked at that at some of the companies I've been with where you actually separate the claims into those people who continue, or who are currently continuing and those people who are no longer with the group and give a higher credibility to those claims where they're still there.

It shouldn't be too hard to work through the math of that, but I could try and dig that out.

From the Floor: You mentioned earlier the goal here is to try to optimize the statistical credibility. If the marketplace is excessive in assigning its credibility, and by optimizing those values you're below that, then it would seem that the consequence of that is that you would be going after the groups that appear in conventional wisdom to be the less healthy or less favorable groups. You have an issue of how you communicate that to your salesforce. If the optimum level is below the market average of credibility, then you're going to have a tendency to be more competitive at groups that have had weaker—poorer experience in the past. That's going to have an impact on how you related to your salesforce.

Mr. Fuhrer: I have a number of comments on that. I have thought about this particular issue. What happens if we're in a situation where the marketplace, for whatever reason, is using a lot more credibility for their rating or, on the reverse, are using a lot less? First thing is that, and I did touch on this in the paper a little bit, if you put the competition in, and you kind of weight your answer by the probability of keeping a group, then it turns out that the optimum credibility actually move a little bit towards the marketplace, but not all the way. I'm not sure what the reason for that is. The only rationale I could come up with was that if they were giving, say, no credibility, then an optimum strategy would be only give credibility on those groups that had poorer-than-average experience. Then you wouldn't get them. You'd give a lot less credibility because you wouldn't want the groups that had poorer experience. Therefore, your rates would be a little bit higher than the marketplace, so you wouldn't get them. Whereas for the good groups, because they're giving them no credibility, you don't have to go all the way. You could still be under the marketplace and still get them. So, if you put that in, it does tend to move in that direction.

The other problem is that the marketplace is incorrect, and you feel you're more correct. There is some safety there because presumably you're going out with the best estimate on all your groups. To the extent the marketplace is different, they're going to lose money. On the other hand, the effect of where the marketplace is too credible, it's going to give you a block of business that on average is worse than everybody else's average. Then you're going to be faced with this big problem of how to set your manual rates. If you just blindly set your manual rates the average of your pool, then that's going to become very uncompetitive and basically wrong. You're going to have to take the leap of faith somehow of deciding that your block of business on the average is worse, therefore, your true manual rates that you should be going out with need to be less than your pool rates. See if you can talk your management into doing that. It would be the right result. I mean if you had

some reasonable knowledge that your credibility had been less than the marketplace's, then you ought to be setting your manual rates less than your actual pool, and vice versa. That might be something that people might miss. If you've been more credible in the marketplace, you're going to get a lot of the better groups. You better load your manual rates. Now you'll have trouble with the salespeople. Of course you probably could put your trend in higher.

Mr. David W. Dickson: Have you thought about modifying your model to look at credibility of claims above certain size levels, for instance, to apply it to stop-loss or to pool claims?

Mr. Fuhrer: We have a lead into the next section of what I was going to talk about. If you only take claims in the second year above a stop-loss point, then you'll get different values for K_1 and K_2 but, most importantly, for K_2. Now, the effect of doing that is to measure the credibility of stop-loss claims based on the experience of all the claims which is exactly what you want to do. If you're rating on strictly manual or strictly demographics for your specific stoploss, that's the way that it's commonly done by a lot of the reinsurers and the big stop-loss carriers. A lot of other companies, though, are doing it strictly on percent of claims. We use 5% for 100,000 or 50,000, whatever it is. I call that the experience method, if those total claims are based on the group. What we have is a method to experience rate specific stoploss where we use a blend of the two methods, where for the very smallest group we'll use mostly the pure sort of demographically adjusted rate, and for the largest groups we'll use mostly a percentage of claims. The effect, though, of using a K_2 that's smaller than K_3 is to mean that even for the very largest groups we still don't get anywhere near 100% credibility. So, even the largest groups end up with about 60–70% blend of the percent of expected claims versus the pure manual rate for a specific stop-loss.

This really requires a hard sell. It seems natural to me, particularly since I've heard some debates about which is more appropriate. The feeling is that whole claims don't tell you much about the claims over 100,000 because they tend to be different ones. I guess that's true. On the other hand, they should tell you something. If you've got an unhealthier block of people, you are more likely to have the big ones. This gives a nice answer, and it really is probably optimum, and if you're competing against the bigger guys who are using only the pure manual rates, then you're more credible. We talked about what that can do. So, in pricing stoploss there is something here.

I just want to quickly go over a couple of things that I've looked at. In the TSA article I talked about how you might set pooling points. These are not necessarily pooling points for retrospective premium but how you're going to pool each group's

claims when you enter into the experience formula. The method to do this is to merely go ahead and do the data set, except pool the claims in the first year and see which level gives you the optimum answer. By optimum, I'm talking about what the minimum least squares is. You can look at the paper for the formula for that. I did this once at one employer, and although the optimum pooling point went up with the size of the group, it didn't really matter that much. If you set a very low pooling point, then you were throwing away a lot of the data. If you set a very high pooling point, you ran the risk of one or two claims sort of dominating things, but the trade-off was roughly even all the way through. So, although the optimum pooling points tended to run about 10% of expected claims, it didn't really matter a whole lot whether it was 5% or 15. So, this is sort of a non-answer, but at least if you need some sort of guide, you could go through this type of analysis. If one of your underwriters asks where should the pooling points be? You could tell them.

One of the things I'd really like to look at is to get three years of data and to expand the formula. There'd be more Ks. We would see what effect two years back has on the coming year, and one year. We could also see what the total credibility of the block of two years would be, as well as the relative credibility of the two years, or even three. That would be very useful for you in the 100% credible groups where the question comes up, how should we weight the years? Everybody uses 60/40. I don't know but 5-3-3-1sounds more like a defense in football. I have not managed to get enough data because you obviously need more groups that are with you that long to really answer those questions. Particularly for the bigger groups I would really like to know the answer. For claim reserves we could also use the method to set the claim reserves where we're actually looking at the reserving factors for a particular group and seeing how credible they were.

Finally, and this you could do with your data, since you're generally in a position of not knowing the claims at the you're actually rating each group, you could take your data and use only the raw claims that have not been completed in the first year or what your completion factor is that you would use. As opposed to what actually came out, you could use those in your data set. That would give you the credibility of the claims as you actually looked at them. I have looked at that, and, of course, the answer is that credibility's a little bit lower.

Ms. Michelle G. Dyke: I work in the individual health insurance, and we have a problem in determining our credibility when we're filing rate increases because we have experience in multiple states. Each state wants to know how much credibility does their state get. Could you use this formula to define a certain level of credibility for a state's experience or is another formula that would work better?

Mr. Fuhrer: This particular formula is only really relevant to the group insurance situation. With that said, I should mention that I did take a peek at the individual problem. There is more credibility to smaller blocks of business probably than most of the insurance regulators would admit. There is a problem, though, that some of these variances are hard to get. So you might actually want to go back to the limited fluctuation credibility that is real old where you're setting rates based on an arbitrary 5%, and that gives you the 100%. You could do partial credibility that way also. I hadn't seen that before where we give partial based on the same criterion. It does give you more than I think most of the regulators would want to use.