RECORD, Volume 31, Number 1*

2005 New Orleans Life Spring Meeting
May 22–24, 2005

Session 6 PD
X-Factor Opinions

Track: Long-Term Care

Moderator: Nancy Westfall Winings

Panelists: Larry M. Gorski
Michael Palace
Erin Colleen Wright

Summary: Since 2000, companies have been required to file X-factor opinions. How has the analysis process evolved? What feedback has been received from regulators? What lessons have been learned? Attendees learn about techniques and approaches that are intended to improve the processes, analyses and reports for X-factor opinions.

MS. NANCY WESTFALL WININGS: The purpose behind developing this session was to take a look at the process over time (as people have now been working with writing X-factor opinions for four to five years), lessons that they've learned and things that might help you in your job on how to improve the process and maybe do your analysis a little differently. We have a very esteemed panel today. We have, first, Michael Palace, who writes the actuarial opinions for X-factors for Transamerica. We also have Erin Wright from Swiss Re, who is intimately involved in doing the actual roll-up-the-sleeves analysis for many of Swiss Re's clients. Finally, we have Larry Gorski, a consultant with Claire Thinking. Larry has a unique perspective as an ex-regulator and also has, on behalf of clients, filed these opinions. We also have a broad audience of individuals who will be able to share with you their learnings and findings of this process.

Note: The charts referred to in the text can be downloaded at: http://handouts.soa.org/conted/cearchive/NewOrleans-May05/006_bk.pdf

MR. MICHAEL PALACE: As I was walking here this morning I passed two clocks, and both of them were set on Standard Time. I took that as a sign that maybe before I launch into the presentation today on X-factors, it may be worth turning the clock back just a little bit to remind those who are old enough to remember what the world was like prior to the adoption of XXX. In particular, many of you probably remember that companies such as ours that sold level term and dabbled in universal life (UL) with secondary guarantees, which in those days meant specified premium products, took advantage of the Standard Valuation Law—or we were accused of taking advantage of the Standard Valuation Law—by coming up with designs such as graded premium whole life, under which companies writing those products were able to avoid setting up any deficiency reserves at all. It was a magical device that eliminated them. This went along for many years until the adoption of XXX with the segmentation.

One of the by-products of the segmentation process is that it eliminated the opportunity to use the tail in order to avoid deficiency reserves. With XXX and segmentation, we no longer have that opportunity, and therefore, for companies who are writing this kind of business, the adoption of XXX was not, in their opinion, a very positive step. So the fact that we have the X-factors, which eliminate or help to eliminate or reduce deficiency reserves for companies who have no deficiency reserves in the first place, could hardly be considered a banner day. On the other hand, for companies who wrote traditional whole life products, my impression is that it was a very happy day, now that they could take advantage of a new approach and therefore significantly reduce, if not eliminate, deficiency reserves. The adoption of XXX and the availability of this X-factor approach therefore put some companies in a better position than they were in before, but for many companies, it created situations that did not previously exist; that is, they did have deficiency reserves or, at a minimum, it put them in the same positions that they were in before. I just want to pour a little bit of cold water on the excitement about X-factors and say that the fact that X-factors are available certainly puts us in a better position than without them, but they were not exactly regarded as a major breakthrough for many companies.

At the risk of belaboring a point that I tend to try to get across in any presentation that I can (whether it's directly related or not), I always try to bring this out: the current environment for reserving on a statutory basis is really not that progressive for companies, especially those in the preferred risk marketplace. Before we go into the X-factor approach, I do want to emphasize, because I do believe that there are many actuaries working out there who may be unaware of this, that on the current statutory reserving for many companies, especially those writing preferred risk products, the level of statutory reserves required, even if you have zero deficiency reserves, is extremely high and excessive.

Since actuaries are supposedly very good at throwing out statistics, I will share some with you. Since the adoption of XXX at our company, we have over $1 billion of statutory reserves on products that we have sold subsequent to the adoption. Of

over $1 billion of statutory reserves, based on what we might consider reasonable GAAP benefit assumptions, our level of GAAP reserves—my corporate actuary is here to make sure that I stay on the straight and narrow and don't deviate and get us into any trouble—is about $300 million. If you then take into account the effect of a deferred acquisition cost offset, we actually get below zero. A gross premium valuation will bring us even below that level to a sort of negative $500 million number.

Clearly, even with all the X-factors, we are still not in a situation where we're setting up reserves that bear any semblance to reality. I think that's a very important point. The reason, again, that I mention this is that there are many regulators and many actuaries who work in companies that are not perhaps affected by this who genuinely feel that the introduction of X-factors and the relief on the deficiency reserve is some kind of major event that actuaries who work for companies such as yours and mine should stand up and cheer about. Just in case anybody thinks I'm standing up and cheering, I'm not.

Now, let's go on to X-factors. The X-factors created by Regulation XXX were adopted for most states on January 1, 2000, although, interestingly, I believe there are still a few states that have not yet adopted Regulation XXX, but I think the opinion would be that everybody is swept in under codification. Under the provision for the X-factors, the appointed actuary has to opine annually that the X-factors meet the requirements of Section 5.B.(3), and, in addition, the appointed actuary's opinion must be supported by an annual report, which was fleshed out and defined by Actuarial Standard of Practice (ASOP) 40.

I did discover recently (maybe you have some similar experience) that some companies are writing traditional whole life products. They may not be fully aware of the fact that under XXX, any company writing any kind of life products is entitled to take advantage of the X-factor approach to reduce their deficiency reserves, if they meet the criteria outlined. Even companies with very traditional products could use the X-factor approach in calculating or reducing their deficiency reserves. This, as I discovered this last week, is not necessarily recognized by a lot of actuaries who work on traditional products.

What are the key provisions? I will review the big six that are actually elaborated and stated in the law (see Palace, slide 3). Very simply, the X-factors must comply with the six tests. They must vary; they can vary by some or all of the policy factors expected to affect mortality experience. They are not to be less than 20 percent. They cannot decrease in successive policy years. Your expected mortality that you are using with the X-factor adjustment still has to be greater than what you actually anticipate. The fifth criterion qualifies this and further states that over the first five years this has to be true. The sixth criterion takes into account any adverse affect on expected mortality and lapsation of any anticipated increase in gross premium. So those are the rules as far as the regulation is concerned.

ASOP 40, which was developed after the adoption of XXX, gave some assistance and guidance and was adopted at the end of December 2000. I was very privileged to serve on the committee that came up with this, although I must say that at this point I don't remember an awful lot about the deliberations. I had to reread it and familiarize myself with some of the finer points for the purposes of the presentation today.

There are some critical issues in the ASOP. You ensure that the X-factors comply with the five-year anticipated mortality test. Companies have to recognize the potential that if they're not careful and the X-factors fail in the future, they may have to increase them, and that suddenly creates the possibility of resultant increases and shocks to surplus. That may be a self-evident truth, but it is certainly worth recognizing that over time, if you're forced to increase your X-factors, it's a one-time, up-front, all-in-one potential impact on your bottom line. You need to give careful consideration to obtaining good quality mortality experience data, you need to demonstrate confidence in the choice of the anticipated mortality and you need to prepare a good supporting actuarial report—again, guidance to actuaries who are about to embark on the sea of X-factors.

The ASOP 40 further defines other criteria for the creation of the X-factor classes, which should generally have similar underwriting or experience characteristics. Mortality experience should be readily available for the defined classes. Anticipated mortality experience should be on a gross (before reinsurance) basis. If the anticipated mortality on reinsured business is considered material, creation of separate X-factor classes should be considered. I'm stating what everybody I'm sure who has anything to do with these opinions has read, but these are the key highlights.

Now, what do you do in practice? Where do we go? I happen to work for an organization that has a lot of business that is affected directly by the need for X-factors. It's a large company. Our exposure over the last 10 years, which we could use to draw on, is over 1.4 trillion policy years, which is approximately 25 percent of the total experience that was used underlying the 2001 CSO. So, yes, I represent a large company, and that brings one aspect to the choice of X-factors. On the other hand, we have a little company that files in New York. There are more people in this room than we have policyholders in New York. Why? Again, that's a very good question, but it's a fact. So I also have to create an opinion for a company that is a very small company, probably smaller than one anybody represents in this room. I have experience on both sides.

In general, in the large company, which has most of our business, you can go beyond the risk class to look at differentiation. You can cut the business any way you want: by size, sex or issue year. If you have enough experience, you may feel going forward that it's a very good idea to try to slice and dice it in as many ways as possible. However, you need to be sure that historical mortality experience is available because when you set up the X-factors, you are under an obligation to be

able to point to experience. In addition, going forward, we took the position that we wanted to make sure that, whatever criteria were selected to slice the business, merging experience would be readily available, that is, that the data would be able to be selected. For example, if you choose to segment your business by zip code, you need to be sure that going forward you can actually study your claims experience by zip code. That was how we decided to set up the classes.

Finally, I don't think it is unreasonable to say that when you set the classes up, it is important to consider the potential impact on the actual deficiency reserves themselves, that is, to recognize that certain groups of policies, because of the way the premiums are constructed, are more prone than others to generate deficiency reserves, and therefore, to give serious consideration to aggregating your business with a view to minimizing the deficiency reserves. That's a reality, and I suspect that most of the actuaries who prepare these X-factor classes take that into consideration. I strongly recommend you do that; otherwise, you may have some very unpleasant and unnecessary surprises when your valuation starts churning out the results.

As an example of what I mean by that, let's assume your mortality experience is equally distributed across all size bands, which is probably not an appropriate assumption, but let's say you're in a company where that is appropriate. It is probably clear that the bigger the premium, the lower your deficiency reserve is going to be for a given level of mortality. Therefore, to some extent, you choose the bands so that you take advantage of the fact that the higher premium business is going to generate fewer deficiencies. If you ignore that, you are effectively not taking advantage of the ability to offset sufficiencies with deficiencies, because a higher premium product can absorb a higher level of a hit to the net premium. Therefore, if you meld everything together, you run the risk of eliminating some of the advantages. You want to give serious consideration to what the deficiency reserves will look like when you set the X-factors. I assume that most people do, and I'm stating the obvious, but I think it's very important. When you choose the classes, try to do it in such a way that legitimately you are appropriately minimizing your deficiency reserves.

The ASOP says that selection is based on anticipated mortality for each class without recognition of mortality improvement beyond valuation date. Again, that's a very important issue. You have to look at where you are today, with no improvement, and then, as the ASOP points out, as uncertainty concerning the level of anticipated mortality increases, provide a margin for conservatism. In regard to consideration in determining anticipated mortality, you look at relevant credible company experience for similar risk classes, other relevant credible company experience, noncredible subgroups (reasonable relationships, etc.) and possible antiselective effects of lapsation or change in the environment.

What do we do? First of all, if you work through the mathematics of how this all works, as time goes on, especially if your X-factors are significantly less than 1,

because the comparison is to a real reserve based on 100 percent of the table, you develop a margin. If your initial choice of X-factors was 50 percent, and later on you are forced to raise it to 60 percent, you have created over time a certain margin in the base reserve that helps cover the impact of raising the percentage over time. So there is a certain element of forgiveness, because the base reserve is still based on the conservative 100 percent standard. However, it is not a good idea to take that into consideration and to wing it.

A large company such as ours can look to its own data and base its choice on its own experience, but the company needs to be very careful in recognizing and assuming that this experience will continue. Therefore it is wise, where it is feasible and practical, certainly, to make sure that you have a margin for adverse deviation in your choice. The aggregation also is going to be determined by the expected distribution of your sales, and you have an opportunity—again, you have to be careful—to assume some mortality improvement up to the valuation date. This, again, is for a large company, where you have years of experience from which you can draw and to which you can point. A small company, on the other hand, has to be more judicious in its choice of these assumptions since it does not necessarily have its own base. Erin will elaborate on that. But even at a large company, you certainly want to be careful and make sure that your experience is very credible before you give 100 percent credibility to it.

To avoid a lot of the issues to do with choice of anticipated mortality, our company is effectively making the statement that our choice of X-factors mirrors our pricing mortality and is also going to be our anticipated mortality. I did have a discussion with an actuary for whom this was not a self-evident truth, and I'd be interested to find out how you actually can set this X-factor. You can base it on a mortality level, if you don't, and then go to great lengths to say that this is not your anticipated mortality. I'm still a little unsure of what that really means, and I'm looking for some clarification on that myself. So I'm glad I'm at this session.

Once you've chosen your X-factors judiciously with all the experience, and you've created the aggregation that you have presumably selected with an eye to where your actual deficiency reserves are going to come out, which I do propose and suggest that a responsible actuary should be doing, you now have to get into tracking experience on blocks differentiated by these separate X-factors. Again, even in a large company such as ours, where we have a high degree of exposure, we do have to use aggregation for credibility because we are tracking experience. This is a decision that we have made. We are going to track experience only on the blocks that are associated with the X-factors. We are kind of saying that the history and the experience on the blocks written prior to the adoption of XXX, which were obviously used in the creation of the initial X-factors, are gone. They're history.

We are now looking at the current block, and we're going to follow its fortunes. Yes, the risk is that the X-factor has to follow the fortunes of that block, but because a large company has that credible experience, it can do that. To me that is the spirit

of what the X-factor is trying to achieve. To constantly refer back to experience of prior blocks when you have credible experience on the X-factor block, I believe, would be a little suspect. On the other hand, if you're like the small company in New York for whom I run the X-factor approach, where we'd be very lucky in the next two decades if we had enough experience, I do refer to the experience emerging under the similar business that is being valued in the larger company. I'm making the case that it's going to mirror that primarily because the underwriting rules are the same. Again, we're not referring back to the historical experience of business unaffected by XXX.

If the initial experience doesn't come out the way you hoped or expected, there is statistical analysis required to ensure that your X-factors do mirror or, at least within the bounds of credibility, are still appropriate. I believe that Larry was the first one who mentioned the Monte Carlo approach and wrote up a paper, and I think that we extracted that from older studies. If the number of deaths in any particular aggregation cell is less than 35, and we wish to test the validity of the X-factors, then we use a Monte Carlo testing approach. I saw in the practice note that there was some discussion. We run 10,000 tests, which I believe is statistically adequate to give you credible results. We do do that. Initially, in the first few years, there were several areas where the aggregate mortality was higher, but since the number of deaths was less, we used the Monte Carlo approach, and in most cases we were okay. On the other hand, where necessary, I did raise X-factors. Those who know me must realize that that is not something I took lightly, but we have to abide by the rules, and we raised them where necessary because they failed the tests. Again, this is a responsibility. We have to be prepared to do this. By the same token, I lowered X-factors in areas where it was clearly appropriate.

What are the special considerations? In our UL business, we have a fair amount of conversions. I'm not giving away any antitrust information when I tell you that our mortality experience on conversions is not quite the same as on our fully underwritten policies. In order to avoid bringing the whole ship down, we separated our conversion business from our fully underwritten blocks. We track them separately, and we also have separate X-factors for the groups where conversions come in.

When we started out we didn't think very much about substandard risks, but now, as we sell a fair amount of substandard business, through our mechanics and even through our mortality studies, we effectively aggregate out substandard, the moment, with the standard risks. Nevertheless, in our actual practice, we use the extra substandard premium to offset the extra mortality that we do use on substandard. We are, I think, appropriately valuing substandard business separately.

Next, let's talk about materiality of X-factors. When I inherited this role from another area, I noticed two things. First of all, they had the X-factors to three

decimals, and, second, we had X-factors of, say, 0.98 on very tiny blocks of business. I cut it down to two decimals, and I simplified the structure, but I also made a decision that on the small blocks of business, if it's 0.98, I'd say to call it 1 and move on. I don't think you will find that your bottom line is going to be severely affected. When a regulator looks at this, if you appear to be too refined and too trite in what you're doing, I think that may trigger an alarm bell. That was just my own feeling. Something that looks a little less precise and less microscopically defined, I think, is actually helpful and may give more confidence. We have an expression we use a lot around our company: We measure it with a micrometer; we mark it with a chalk; then we cut it with an ax. Some of that applies here in terms of how these X-factors are defined.

Another special consideration, which for us is a very big deal, is the sunset provision on the estate tax. We do wonder a little about that. We have a fair amount of second-to-die or joint survivor business. But we do actually use the Frasierization; we use the formulas in the calculation, and we apply the X-factor embedded just the same way as in our base policy. Some of this is very complex, and applying the X-factor layered on top of the base valuation is just another layer of complexity on these second-to-die products.

The actuarial opinion is where rubber meets the road. The opinion, first of all, is very important. You must define your qualifications for the job. You also have to confirm familiarity with the model regulation, ASOP 40, which means you have to, I think, read it very carefully at least once a year before you embark on this. The opinion should indicate, other than the valuation date, whether the company's X-factors meet the requirements of the model regulation. The opinion should reference, where appropriate, the supporting report. The opinion just says that you do the right thing.

Then you have the report itself. You have to define the purpose of the report. More importantly, you provide a description of the plans; the in-force statistics, including the reserve amounts; mortality studies for covered plans; and the schedule of X-factors, noting in particular whether any have changed from prior reports. I tried for completeness here, but to some extent I wonder whether it's better to have too much or too little in some of these reports.

MS. ERIN WRIGHT: I work in the research and development area at Swiss Re Life and Health North America. I'm going to focus on a more technical aspect of X-factors. I want to give you a little more background not only on XXX, but also, as a reinsurer, on what our experience is with X-factor analysis. I'm going to expand a little on what Michael talked about in his presentation on prospective analysis, but the bulk of my presentation will be on the retrospective X-factor analysis.

Under Regulation XXX and ASOP 40, you have the option of using a percentage of the underlying CSO table to calculate your deficiency reserves, but in using this option, you are required as the appointed actuary to opine that the X-factors follow

six tests. That's the prospective part. But then you also have to do a periodic review against the emerging experience, so that's the retrospective part. As a reinsurer, we help clients with both the prospective and retrospective analysis. At times, our pricing area helps our direct insurers with creating an anticipated mortality assumption and then with calculating X-factors as well. In my area, research and development, we help certain clients who don't have the resources with analyzing their emerging experience for this periodic review, which is the retrospective part.

To review Michael's presentation, prospectively you need to create your anticipated mortality assumption, and then you need to calculate your X-factors using the six tests. The definition of X-factor class is something Michael touched on as well. Do you aggregate, or do you not aggregate? We call the nonaggregation a standard method of defining X-factor class. Where it's varying, the X-factors are varying by everything by which the mortality varies. Then Michael touched on the aggregated or grouped method, where your X-factors do not vary by all of those factors by which the mortality varies.

There was a survey done in 2002 on XXX. One of the questions had to do with X-factor classes. Sixteen percent of respondents said that they vary by five factors in their X-factors, 32 percent of companies used four factors and then it went down to 27 percent who used only a single factor. It was spread over the spectrum. I want to take a poll of the audience. For those of you who are aware, which of you use the standard method, or do you have X-factors that vary by everything by which your mortality varies? Not very many here. How about grouped—those of you using grouping or aggregation in your X-factors? Definitely a few more there.

We've already talked about the six tests (see Wright, slide 5), so I won't elaborate too much. I just talked about the first test that says you can vary by all or some of the factors that are expected to affect mortality experience. I'm going to be talking about Tests 2 and 3 (the 20 percent and the nondecreasing rule) as well.

I'm going to elaborate a bit more on the standard method. Again, Test 1 says that you can vary by all of the factors, if you care to do that. In the standard method, as Michael pointed out, you can differentiate by gender band, etc. You can aggregate, or you can do it by all of the factors affecting. The advantage of using the standard method, where you're essentially varying at the cell level, is that if you vary by every factor that impacts mortality, then there's no risk of distribution shifting underneath if you grouped your X-factors. Disadvantages are that when you bring in Tests 2 and 3, you're likely to hit that 20 percent floor more often at the cell level, like superpreferred, nontobacco, or something like that, and then Test 3, the nondecreasing rule, can create some margin of anticipated mortality as well, especially if you're hitting that rule at a very early duration. So, for this margin with two tests built in, it's not bad to have conservatism; it can definitely be an advantage if you have been too optimistic when you're setting your initial mortality assumption, but, of course, it impacts your deficiency reserves.

Again, in aggregating or grouping, your X-factors don't vary by as many factors as your anticipated mortality. For instance, maybe your anticipated mortality varies by issue age, but you don't want your X-factor to vary by issue age. I've seen companies do this before. Instead of doing a weighted average of your X-factor, you actually want to weight together your anticipated mortality using your age distribution, and then calculate your X-factor using the six tests. That's an important point to make. There are some advantages of using an aggregated or group method. Michael covered this quite a bit, so I won't say too much, but I did want to point out, again, Tests 2 and 3. When you aggregate, there are times, when aggregating over age, for instance, that you're going to be hitting the 20 percent floor less, and then you also may have the nondecreasing rule not even kick into effect or, if it does, kick in in a later duration. So the margin is not as excessive as it is when you're doing the standard method.

A disadvantage would be that if you have a young block or an open block, your distributions could shift over time. There's a changing demography in the block. As that block matures, the distributions will stabilize, and stable distributions are what you want. But if you use this young block, which probably many people did back in 2001 when they created X-factors for the first time, and if you look at this over time, then it may be necessary to change your X-factors as distributions shift.

Like I said, I want to concentrate on the retrospective X-factor analysis, which is more what my area does. To reiterate, retrospective analysis is the periodic review of emerging mortality experience to assess the adequacy of the anticipated mortality used to calculate your X-factors. So how do you review or test the adequacy? You want look at actual-to-expected ratios both by number and by amount. We also look at statistical hypothesis tests. Monte Carlo was brought up already. At Swiss Re, we're big fans of the Panjer Method. I'm going to talk about those methods in a little more detail. I'm also going to address data issues, the analysis itself and then resetting X-factors if that becomes necessary.

In a Monte Carlo, essentially you're simulating an aggregate claims distribution through a number of trials or scenarios. The number "10,000" was brought up, but I think that people still ask this question: How many scenarios? We use 10,000 in our final analysis run. That number is most commonly used because it stabilizes the distribution. However, there is the issue of runtime, so you do want to consider that as well. When we're doing our analyses, we're making code changes based on several runs cleaning the data. When we do this, often we'll reduce the number of Monte Carlo trials to maybe 1,000 until we get all the data cleaned up, and then we can increase it back up to 10,000 for the final run.

Speaking of runtime, I'm going to talk about the Panjer method a little. The Panjer Method, if you haven't heard about it, is based on Harry Panjer's 1980 article in the SOA's *Transactions*. It's called "The Aggregate Claims Distribution and Stop-Loss Reinsurance." It uses expected X-factor mortality, exposures, number of policies and face amounts to calculate the aggregate claims distribution, and essentially

what it does is calculate percentiles that represent the aggregate claims distribution. Monte Carlo takes up to three to four times as long with 10,000 trials. So the claims distributions for these two methods, Panjer and Monte Carlo, do converge as the number of trials increases. Again, we prefer Panjer because of the time economy, but I would say that—Larry can probably address this more—Monte Carlo is still used by and large because of its familiarity in the industry.

Next I'd like to address data. Since the point of the analysis is to look at claims experience, I'm going to start there. We assume that pending claims will be paid, and so we include them in our analysis. This is conservative, but if it happens that failures occur in the final result, then this conservatism can be taken into account with the average claims denial rate when trying to dig into the results further. It's appropriate to make an adjustment for incurred-but-not-reported (IBNR) claims. There are a couple of ways we handle this, and I'd be interested to hear if anyone does this differently. One way we handle it is to check out the data, see when the terminations stop stabilizing and cut off the valuation date at that point. Another way of handling it is that we get the data sent in as of a certain termination date, and then we ask for a claims update a month or two after that to make sure that all of the claims lag has been addressed.

Regarding claims as a result of extraordinary events such as 9/11, I would say that we think it's appropriate to exclude those claims from the final report. However, it's probably recommended to do the report both ways, with and without the claims. Then if you do exclude these claims from the final results, it's probably a good idea to make note in the opinion that those have been considered.

Sometimes we don't have the actual date of death. This can skew the results a little. If you can produce the results with actual date of death, that's recommended, but sometimes you have to work with the report and pay to make adjustments to those as necessary. We ask for cause of death. In the event that there are failures or things that we want to look into more deeply, then cause-of-death information can be useful to look into underwriting and claims contention practices of the company.

Michael talked a little about pre-2000 issues. Actually early on, after 2000, we spoke to some regulators about their opinions on this, and, for credibility issues, we did find that they were recommending using pre-2000 issues. As your blocks age, is it appropriate to do that? I would say that in the event that you do not have any credibility even by aggregation, then pre-2000 issues of a similar underwriting would be appropriate to combine with your blocks.

Data quality is a big issue. There are some questions that you want to consider. We have to ask these questions of our clients because they send in their data to us. You may already know those questions because you're the one creating the data. How are exposures calculated for new issues? For terminations? For joint life? Also, can you identify or map all of the variables that you need that are affecting the X-

factors? It's important to be able to do those things. I would say that "garbage in, garbage out" is a good adage for data quality. A majority of the time that we spend in our analyses is data related. We're fixing it, cleaning it, rerunning it or making code changes for it. It's a good idea to streamline a process to create a data set that is reliable and a way of verifying its accuracy, because it's going to save you a lot of time in the future. You probably already know that.

Let's move on to the analysis itself. I'm going to talk about the actual-to-expected (A/E) analysis, confidence interval analysis, credibility, aggregate versus individual X-factor results and then trends. To give you an idea of a report that we put out, this chart (see Wright, slide 13) is a univariant analysis or summary for ABC Company (not a real company). As you can see on the top line, these are the results in aggregate, but then we have a breakdown of results at an X-factor class level as well. Then we have both by number and amount expected, exposures and actual claims. We have also the A/E ratios by number and amount and the Panjer percentiles.

Why is A/E analysis important when ASOP 40 says that you only have to do the analysis by amount to look at both? I have a couple of examples. The age-60-and-up group is the first example (see Wright, slide 14). You'll notice that the A/E ratio by amount is far less than the A/E ratio by number. What's going on here? Well, the large claims experience has been very good for this company. Ignore the fact that they only have one claim, please. Pretend that they have 100 claims. So what can you expect? You can expect that that's not going to go on forever. As they build credibility, those large claims results are going to stabilize. That's going to catch up with them. The point here of looking at the A/E by number is to ask, is this a healthy number? This is not terribly healthy, so you want to focus on this as sort of your ceiling. If you have an unhealthy A/E by number, then that's a warning sign. You need to keep an eye on that. Another culprit we've seen many times is excessively large claims. In the 250 band, the A/E by amount is much higher than A/E by number, which is probably something that you see a little more often. You can kind of filter out that effect through capping in your analysis, if you want to get rid of that large claims effect.

Regarding confidence interval analysis and statistical hypothesis testing, the null hypothesis is that the X-factors and the anticipated mortality are adequate. You set a level of significance, say 5 percent, and then you can accept or reject this hypothesis based on the results of the test. In our case here (see Wright, slide 15), we have a 5 percent level of significance. We have failures when the A/Es exceed the 95th percentile or, as you can see, we have the Panjer percentile laid out, but it's at the 98th and 97th percentiles. So that's a failure. We have failures both by number and by amount. What if results fall below the 5th percentile, for instance, on a two-sided test? That could indicate that you have some excessive margin there, and maybe your X-factors could actually be reduced, which is something you don't hear about as often. That's something to consider, but I would say that from a regulatory perspective, they're more concerned about the above-95th percentile.

I'm going to talk about credibility. What is the rule of thumb for credibility? We start assigning credibility to results at 100 claims, approximately. Let's say that your test fails in aggregate, but you have only 20 claims. What can you do? There's not enough credibility there. We recommend periodically monitoring the results until that experience emerges to make it credible. Another idea is to combine similar blocks. Michael was talking about New York versus non–New York and similar blocks. Maybe they have different X-factors, but you can bolster the credibility by combining the two. Another thing you can do is add pre-2000 issues.

We talked about the aggregate level. What about passing at the aggregate level but then failing at an X-factor class level or an individual X-factor class level? What happens then? We're five years along now, but at first, very few companies had enough claims to even be credible and in aggregate the first few years. But we've seen some companies who have over 100 claims not only in aggregate but at lower levels, say the plan level. You need to look at both aggregate and individual X-factor classes for that reason. If there are failures at the lower levels, and the results are credible, then you need to dig into that further. You need to analyze for the large claims effect. Multiple claims per life can be pesky as well. It's good to try to eliminate those effects, not necessarily for the final results, but you're going to be looking at several iterations, and one thing to do is tapping to get rid of the large claims effect so you can see what that would look like without those.

I am going to bring up cause of death again. We found that when results either don't look like we think they're going to look or when they're failures, and they don't seem to make sense, we want to dig further into each individual claim and look at some cause-of-death information. Maybe this is going to uncover some things about the underwriting and the claims contention policies. What if there's a very high claims denial rate? Does this mean that they're just aggressive in their claims investigation, or does it mean that the underwriting was poor in the first place, or maybe both? You may unearth some things that would indicate there needs to be some internal changes. If internal changes can't be made, maybe you need to raise your X-factors. Failing results could also be random fluctuation. There are a lot of possibilities there.

What if your tests reveal that X-factors and the anticipated mortality are not adequate? What if further investigation suggests that you need to reset your X-factors? What do you do then? There could be minor changes, or there could be major changes. As an example of a minor change, maybe you're just changing a single class; you're not changing your X-factors or mortality across the board. Maybe when you created your X-factors, Company ABC had a single preferred nontobacco class. They assumed that they were going to be competing against other companies that had a single nonpreferred nontobacco class. A couple of years later, people start introducing superpreferred, but ABC still hasn't, so they're getting those risks who couldn't qualify for superpreferred. They're coming over to ABC now, and they're getting those bottom-half risks, if you will. What's happening

when they look at the experience? It's not emerging like they expected. It looks like maybe they need to change their X-factors based on this new experience, this experience that looks different from what they expected. That could be a minor change. That would be a change to a single X-factor class as opposed to all of the classes that weren't affected. Another example is that maybe your anticipated mortality assumption was just too aggressive. Maybe you need to make a minor tweak across the board to all X-factor classes.

An example of a major overhaul would be that maybe there was a miscalculation in the relationships between preferred and residual. All of your X-factors are affected by this, and you need to start from scratch, essentially. I've focused on when you need to increase, because that's obviously of more import to everyone here, but we have seen companies that have excessive margin and enough credibility to suggest lowering the X-factors. So that happens, too. Ultimately, resetting your X-factors, when and how much, is going to be dictated by what regulators expect.

MR. LARRY M. GORSKI: When I was asked to be a part of this panel, I was pretty excited about it because, as Michael mentioned, I was a regulator for many years. I was heavily involved in the development of the version of Regulation XXX that people are now utilizing, and I was very curious to see what the regulators of today think about when they deal with Regulation XXX. First, I will make a few comments about Regulation XXX and why there may be some changes coming in the future, at least with respect to X-factors. While the process of developing the current version of Regulation XXX was very long and painful (I think if you count up all the years, it was probably 15 years' worth of work to one degree or another), the actual amount of time spent on the X-factor component of the final version of XXX was short. I think everyone was kind of tired, and this seemed to be a way to bring together a couple of sides to the issue. Everyone bought into it, and there wasn't as much thought to the X-factor component as there has been with other NAIC regulations and model laws.

This may come as a surprise, but most of the NAIC laws and regulations are really recommendations from the American Academy of Actuaries that get adopted by the NAIC. In this particular case, this was not. This was an industry document that was reviewed by the NAIC and then adopted by the NAIC. Because it wasn't an Academy recommendation, some of the normal things that you see in NAIC laws and regulations don't appear here. For instance, the language concerning the opinion isn't standardized. The X-factor report is left open to the judgment of the actuary supported through the ASOP. The kinds of things that you normally associate with regulatory requirements are not part of the process.

Since 2000 when this was adopted, regulators have been heavily involved in another project, the C-3 Phase II project. I'm not going to get into any details of that now, but that project, I think, has generated a lot of questions in regulators' minds in four key areas that do have a bearing on X-factors: data quality, aggregation, sensitivity testing and credibility. It wouldn't surprise me if regulators

at some point in time reopen the X-factor component of XXX and, using their current experience with this other project, somehow try to translate it over to the X-factor analysis. I'll try to be more specific on those four points as I go through my presentation. What I'll be talking about is the regulatory view and not my view. I obtained the regulatory view through a survey, so I'll give a compilation of results. I will give my interpretation of everything I heard from the regulators on this topic.

The survey wasn't a scientific survey by any means. I sent it to 12 states, and I received responses from nine states. The questions were designed to be neutral and not biased toward a specific answer. I wanted to get the regulatory view and not lead people to an answer that I would have given if I had still been a regulator. The survey was a multistep survey: I sent the survey out; I got back a few responses; I took the responses; I incorporated the responses into a draft of this presentation; I sent out the survey again to everyone, along with a draft PowerPoint presentation, so that they could see how their responses were going to be used, and, for the people who did respond, they had a chance to see the responses from other regulators. It was sort of an iterative process. The one thing I note is that with some of the later respondents, since they had the advantage of responses from the other regulators, their responses may have been a bit more sophisticated, let's say. They got at a few issues that maybe weren't addressed in the earlier responses.

I had to assure all the regulators that the responses would be held confidential and that I would be true to the response. So the ways I tried to satisfy both of those requirements were, first, not to identify anyone by name or state, and, second, when I have a regulatory response, I have it in quotes. The responses that I'll be giving you are actually what the regulator said in response to survey questions. As I look at this first chart (see Gorski, page 6, slide 1), I notice that the last response in the column on the right doesn't have quotes. It actually is a quote from a regulator. What I tried to do for each one of the survey questions was to take the responses, break them into two camps (right column, left column) and give a consensus view, so to speak.

The first question was: Are you satisfied with the content of X-factor opinions? If not, what problems have you encountered? The overall view was that regulators are satisfied with the X-factor opinions. On the other hand, on the negative side (the right side), at least a few regulators expressed the view that the structure of the opinions and content of the opinions vary from company to company, and that is maybe a little distracting to the regulators. They'd like to have more consistency and more standardization of the opinions. That may be one thing that regulators do look at in the next go-round, if there's going to be a next go-round. I'm just trying to get the gist of what regulators are saying to these questions.

The next question (see Gorski, page 6, slide 2) was: Are you satisfied with the quality of documentation in support of the opinion? Again, in general, the left

column does express a high level of satisfaction. However, there are a few points here that I do want to point out. The fact that regulators are maybe not expressing concerns over the opinions or reports of memorandums in support of the opinion may mean that they may not be reviewing them every year. For instance, consider the response, "We no longer routinely review the X-Factor Report because we were mostly satisfied with them initially." I'm not questioning that person's response, but the initial review of X-factor opinions and reports may have been performed with a lower level of understanding of some of the issues that actuaries face when they produce the reports. Although someone was satisfied in year one or year two, given all of the experiences that have taken place since then, a review today may result in a different response from a regulator.

Regarding the response, "…we tend to only review supporting documentation for 'X' factor opinions as part of the financial examination process," that means the reports may get reviewed once every four or five years, and so there are probably many cases where X-factor reports for a company haven't been reviewed yet. So don't take a lack of response as meaning that there are no problems out there with the opinions or reports. On the right side, one regulator did express a negative comment about the quality of documentation, saying, "…however, the actual report could be expanded, especially in the area addressing data quality."

Of all the questions (I had only five or six questions in the survey), the responses to the question dealing with data quality were maybe the most dissatisfying to me. It led me to believe that regulators are not considering the issues of data quality so far. I don't think any regulators are here, but they may download the material for the session or even buy the CD. There are some good comments concerning data quality issues. I was disappointed with the responses, but that may change over time.

The question dealing with data quality (see Gorski, page 6, slide 3) was: Does the actuary address the issue of the quality of data used in the analysis of X-factors? I even tried to lead people to consider ASOP 23 in their response, but I didn't want to go any further than that. I didn't want to put in specific points that are worthy of consideration from a data quality perspective, but I did point to ASOP 23. Again, I was disappointed with the responses shown on the left side. "Some companies do send in the actuarial report which goes into considerable detail about the data, the experience, and the type of testing that is done." That seemed to be a positive response. On the right side, the response is, "No, except for sometimes mentioning credibility issues if there is a claim." Well, I'm not sure that I'd put a credibility issue concerning a large claim in the category of data quality. It may be part of that. There are other data quality issues that need consideration. One respondent said, "No, not completely." My feeling is that data quality is not a big issue with regulators because they probably haven't thought about it.

When I was still with Illinois, in the first couple of years of implementation of X-factors, we would actually try to rerun the Monte Carlo simulations that companies

were doing. Fortunately or unfortunately, Illinois has several medium-sized companies that write term business and utilize X-factors to reduce deficiency reserves. So we worked with about six or seven companies over the two to three years I was there when this was already in play. Companies would provide us with a data file, and I wrote a SAS program that did a Monte Carlo simulation. We would focus in on smaller blocks of business so we could test to see how many simulations or iterations one would need to feel comfortable with the results of a Monte Carlo simulation. In the process of doing that, I and my associate, who was heavily involved in this also, really became familiar with the kinds of data quality issues that could arise in terms of determining how one counts exposures, missing data, etc. It seems that maybe as regulators begin to refocus their attention on X-factors, which may come within the project or the UL-with-secondary-guarantee project, they may want to start doing some of their own testing and analysis. Once you do that, you bump up against data quality issues, and you have to address them.

Another question (see Gorski, page 7, slide 1) was: Does the actuary utilize Monte Carlo or other methods to evaluate the appropriateness of prior "X" factors? Are you satisfied with the quality of the work performed by the actuary in this regard? In looking at the results, it's clear that a broad range of methods is being used. I would say that, in general, companies are using a Monte Carlo approach. Since I left Illinois, I have done some part-time consulting work with Claire Thinking, and we do some work for the State of New York and some New York companies. From my own experience, the vast majority of companies who do retrospective testing of X-factors do use Monte Carlo studies.

The last response gets to the simulation question, and the person there was questioning the number of simulations. It looks like the company did about 1,000, and the regulator suggested multiplying that by 5 or 10 and getting better results. I thought the second response was interesting in that at least one regulator is referring to the Actuarial Standard of Practice. I was wondering about that. I didn't want to come out and directly ask, "Regulator, do you look at the Actuarial Standard of Practice in your review?" But at least one person volunteered the fact that he or she does, which made me feel good. That document is getting attention by the regulators. Now I have to say—I may break some confidentiality here—that that response did come from the Illinois department, so I feel good about that.

The next question (see Gorski, page 7, slide 2) was: Does the actuary address the issue of mortality improvement beyond the valuation date in an acceptable manner? This is another issue. I didn't put it in my list of issues from C-3 Phase II, but I think it's an issue to which regulators should pay more attention. In many cases, the direct writer is utilizing the services of a reinsurer for his X-factor analysis, and the whole issue of the anticipating mortality assumption—how that anticipated mortality assumption was developed, does it include any mortality improvement in that anticipated mortality assumption—gets lost in the whole process. Going from the reinsurer to the direct writer to the report in support of the opinion, in the cases with which I'm familiar, there really isn't a good direct

response to that question. Of course, everyone is going to say that they're not including mortality improvement beyond the valuation date, but there's very little, if any, discussion as to the support for that statement. I think that's one area that regulators may start looking at because, again, this is one of those issues that did get a lot of attention in the C-3 Phase II report.

Another question (see Gorski, page 7, slide 3) was: Do you have specific concerns with actuarial work in conjunction with "X" factor opinions? The negative side (the right column) deals with the data quality issue and standardization. I think that regulators with an increased knowledge base, if and when they begin digging into these reports on a more frequent basis, may be asking questions in those areas.

This is an interesting comment that I received from one of the regulators: "We found a company today that reduced their deficiencies reserves by 95 percent by using factors . . . I will be reviewing and testing the X-factors . . . I don't think it needs to be in the public certification, but . . . it would be helpful if it were in the report, which we treat as confidential." When I got that e-mail response from the regulator, he quickly called me. The regulator clarified that the decision to review and test the factors was based on the dollar amount of the reduction in deficiency reserves and not the percentage change. This interesting comment leads to an observation concerning the memorandum (not the opinion of certification, but the memorandum).

I think it would be very useful if there were a disclosure of the impact of using X-factors on deficiency reserves, and it may be useful to provide some sensitivity analysis of the choice of X-factors. If you increase X-factors by 5 or 10 percent, what will that do to its impact on deficiency reserves? Because the whole issue of the opinion and memorandum really is not part of the regulatory requirements, it probably didn't get much thought or attention to the need for disclosure of the impact in sensitivity testing. But in the C-3 Phase II project, in which many regulators are involved, from both the Life & Health Actuarial Task Force and from the Risk-Based Capital Task Force, this whole notion of disclosure and sensitivity testing is taking on a much more significant component of regulatory thinking. Based on this comment and what has taken place, it would not surprise me if regulators start asking for that kind of information.

My interpretation of the survey results is that generally all respondents were satisfied. However, three respondents represent states with very few insurers that use X-factors. Because of this limitation and the limitations that follow from the fact that the reports may not be reviewed each and every year, you can't interpret the results of this survey as being widespread acceptance of the work being done by actuaries. Monte Carlo analysis of X-factors appears to be a norm, but some regulators are not completely satisfied with the implementation of the methodology. Lack of feedback concerning the X-factor report may be a function of the review part of the triennial examination process. Again, the level of review may have been

very high in year-end 2000 and year-end 2001. People felt satisfied and then maybe cut back on their reviews somewhat.

Regulatory issues with the level of review of the quality of data and standardization of the opinion memorandum may start to emerge. I think we've talked about that. I think X-factor review has sort of taken second stage for regulators because their attention is focused on the Actuarial Guideline 38 (the UL with secondary guarantees) and C-3 Phase II concerns. What I see happening is that a blending together of the thoughts that have emerged in the C-3 Phase II is going to surface in the UL with secondary guarantee, and the extent is still relevant. They may emerge with X-factor analysis for traditional term products.

At least one state uses ASOP 40 in its review of the X-factor opinion. When I was a regulator and I was called on the phone to answer a few questions or given a survey on a particular topic, my immediate reaction was, "People are interested in this. Why? Have I been missing something in my review?" So I didn't want the survey to somehow scare regulators into thinking that they need to do more work in this area. I tried to be as neutral as possible, but, regulatory nature being the way it is, I'm pretty sure that participation in the survey may cause more attention to be shifted to X-factors.

There's one issue that's not in my presentation that I want to get to before I get to the few comments on peer review. Michael and Erin talked about the anticipated mortality assumption. One reason why regulators are not satisfied with the Monte Carlo testing of prior X-factors is that, in general, they're applying it to small- and medium-sized companies, and its ability to give definitive answers, because of confidence level issues, is fairly limited. I think every regulator likes to see companies failing these tests so that they can increase X-factors, and you're not going to see that happen. So they've put a Monte Carlo–type testing for smaller to intermediate-sized companies on the back burner. In trying to address that issue, both from a practicing standpoint and from a talking-to-my-former-friends standpoint, I suggested that they begin to utilize the SOA surveys that have been conducted over the last couple of years.

The preferred mortality survey of reinsurers and the more recent preferred mortality study of direct writers both provide an awful lot of information on companies' expected mortality based on stylized or idealized underwriting rules. So what I've done and what I've been suggesting that other regulators think about doing is, when they get to a company where the review of X-factors is important, they first review the company's underwriting rules, map those underwriting rules to one of the idealized underwriting guidelines in the surveys to which I refer and then take a look where the company's anticipated mortality assumption falls relative to the range of responses to the survey. You'll get some feeling as to whether a company is being aggressive in its anticipated mortality assumption or middle-of-the-road or conservative. So I've been suggesting to regulators that they think of that, not in lieu of the Monte Carlo studies or historical studies of X-factors, but as a

complement to that and as another thing to take into consideration when reviewing X-factors.

The last part of my presentation deals with a topic that's coming out of the C-3 Phase II project. It deals with the notion of peer review. One of the things that is emerging in that realm of C-3 Phase II is that regulators are not comfortable with simply relying on the valuation actuary. There's serious consideration being given to a regulatory-required independent peer review of the work done by the actuary for C-3 Phase II. The project is being led by the American Academy of Actuaries Standard Valuation Law 2 Work Group. It's an attempt to give additional confidence to the regulator through this regulatory-required independent review that the work of the actuary is acceptable.

It's possible that at some point in time the same concept will migrate over to work being done on the UL secondary guarantee. It may migrate over to X-factor analysis; it may migrate to a much broader area. The thinking of the Academy group is that it's not limited to C-3 Phase II work. It's broad in scope. It may be tested for the first time in that area, but there is at least a chance that at some point in time the scope of actuaries' work in areas like this, where judgment plays a big role, will be broadened to include the X-factor opinion memorandum.

MS. STEPHANIE J. KOCH: My questions are a little technical, and I understand that aggregation could be one way to address them. The first question is, in a situation where your evaluation of mortality would recommend an X-factor in excess of 1, is there a recommended approach? Should you use an X-factor greater than 1, or should you not use X-factors?

MS. WRIGHT: I'll try to answer your question, but I may pass it over to the other panelists. I will tell you that we have seen clients who use X-factors greater than 1. We've been going around this last cycle of analyses. There were clients who had blocks where the X-factor was simply 1, but then other blocks where it was not, and there was some vagueness about the regulation to us. Do you have to include those X-factors or those blocks where the X-factor is 1 or greater in the analysis? If any X-factor is below 1 in the entire block, then all of the block must be included. I would say that because the entire block is included, for that part of the block where the X-factor would exceed 1, I think that if it's part of the analysis, you would definitely want to make it greater than 1. Does that answer your question?

MS. KOCH: Yes, thanks. I had a second question. On the issue of credibility, when you're looking at a block that is either all juvenile or largely juvenile, you may have a lot of difficulty getting to 35 claims, which was sort of suggested as your standard. Is there any rule of thumb? Is there any recommended approach?

MR. GORSKI: The issue that you're raising is one of the issues that regulators talked about and why they tended to maybe push aside Monte Carlo testing, and that's why I talked about the other approach. See what other companies are doing.

It's sort of the peer review type of approach. I can't recall the issue age bands that are used in the preferred mortality studies to which I referred, but to the extent that there's any information about how other companies are dealing with the issue of anticipated mortality in these age bands where there's very little experience, that's what I would suggest.