RECORD, Volume 31, No. 2*

New Orleans Health/Pension Spring Meeting
June 15–17, 2005

Session 3PD
Predictive Modeling Applications

Track: Health

Moderator: Lori Weyuker

Panelists: Jim Minnich†
Lori Weyuker

Summary: Can we improve upon traditional methods of risk measurement and risk selection? An affirmative answer could mean a competitive edge for your organization and a more efficient health insurance market in general. Panelists discuss potential applications of predictive modeling, including individual underwriting, group rating and provider reimbursement. They focus on ways in which predictive modeling can be used to improve risk selection, experience rating and provider reimbursement.

MS. LORI WEYUKER: My name is Lori Weyuker, and this is Jim Minnich, the national health care practice leader of underwriting services at Reden & Anders in Minneapolis. In addition, he's had a long career of heavy experience in underwriting at Prudential, Medica and United Health Care and has consulted with many health insurance companies and HMOs. Jim is going begin our presentation, speaking on predictive modeling applications specifically relating to underwriting,

I'm an independent consultant in the San Francisco area. I've been in the health care area for over 10 years and used to be a pension actuary. I've worked for both national HMOs and national health insurance companies. My presentation on predictive modeling applications will focus specifically on data and how it impacts predictive modeling applications, as well as how to choose a good predictive model. Here's Jim Minnich.

---

†Jim Minnich, not a member of the sponsoring organizations, is the national practice leader of underwriting services at Reden & Anders Ltd. in Minneapolis, Minnesota.

Note: The chart(s) referred to in the text can be downloaded at http://handouts.soa.org/conted/cearchive/neworleans-june05/003_weyuker_new.pdf.

MR. JIM MINNICH: It's a pleasure to share my thoughts with you on some of the changes that I've observed happening in the underwriting process and how today's technology is improving the way we can underwrite. I thought I'd start with some of the current underwriting challenges that I've observed. Over the past six to 12 months, we've seen a significant increase in price competition. The for-profit companies have been experiencing record earnings. The nonprofit BlueCross organizations have excess surplus in most states. Wall Street is pressuring the for-profits to get more aggressive and to grow membership. State regulators are pressuring the BlueCross organizations to reduce surplus. Employers are price shopping and negotiating for better rates. We're seeing more and more consumer-driven and self-insured plans. All of this puts pressure on the underwriting process.

Three or four years ago, I was talking to a lot of companies about predictive modeling, but the time wasn't right. I think the underwriting cycle that we were in at the time was so focused on driving earnings that underwriting could be a little sloppy. They were allowed to get by with two or three points of margin in their pricing formulas because all of the competition was focused on growing profits. I would suggest to you that that cycle has changed. We're now seeing a need to be more precise in the way we underwrite. If your company isn't precise, there will be a competitor that will take advantage of that lack of precision. It will be hard for you to renew your better risk groups, and it will be hard for you to grow with good risk groups.

In the new business section, most carriers I work with still use a manual process, resulting in incomplete information that makes the new business rating process less effective and efficient than it could be. In a small group renewal process, carriers are using traditional approaches to renew their business, which leaves them in danger of losing some of their better risk groups. These are generalizations, but they're observations from clients that I've worked with. The current underwriting process in many underwriting shops is paper-intensive. It's manual, inefficient, a poor use of an underwriter's time, slow and administratively expensive.

When I was at United Health Care in the late 1990s, I was the chief of underwriting operations. I did focus groups of all of our underwriters, with eight to 10 underwriters at a time. The No. 1 suggestion for how we could help them be more effective was to automate more. The feedback I got indicated that over half of an underwriter's time was spent gathering data, validating data and inputting data. We weren't using underwriters as effectively as we could. A lot of underwriting time was spent doing non-underwriting tasks. Of equal concern is that decisions being made by underwriting are often made with less-than-complete information or, sometimes, inaccurate information. Across the country, health plans that I work with by and large have fairly inadequate management reporting. A manual and paper-driven process makes it difficult to get current, meaningful, measurable data to help managers to manage their block of business. All of this leads to potentially compromised results.

What are the goals of introducing predictive modeling into an underwriting process? More automation, better information, freeing up your underwriters' time so that they can focus on underwriting, and allowing them to do a better job of matching revenue to expenses will all hopefully allow an improvement in your expense ratios or loss ratios. Being able to use a tool that is automated and electronic will enhance your management reporting and lead to what we're all desiring—profitable growth.

What is a predictive modeling tool? It is a software tool that predicts the relative health care cost for the next year for each member. It is also a management tool that allows health plans to stratify members by risk and cost levels, more accurately estimating future costs and identifying potential risks earlier in the process. Why do we want to use predictive models? This year, 6 percent of the members drove 58 percent of the cost. Most underwriting processes use historical cost as the means to identify better risk groups from higher risk groups. One of the things that we observed is that last year those same 6 percent only drove 23 percent of the cost, and next year will only drive 24 percent of the cost. Stated differently, if your underwriting process takes a look at those individuals who this year have exceeded $10,000 in claims, and you have a process for a medical underwriter to review the details of that member, and that's the primary basis for you making your renewal risk decisions, you're missing a lot of risk. If we focus just on the 6 percent who are driving most of the cost this year, you're going to miss a big chunk of the cost drivers for next year.

The applications for predictive modeling are new business underwriting which is a relatively new application in the past couple years; renewal underwriting; proactive medical management; actuarial analysis; and employer reporting. I'll focus on the first two, new business underwriting and renewal underwriting. Again, I make some generalizations, but they're observations from having worked with a number of health plans.

In the new business small group underwriting process, what are some of the concerns? Most carriers ask employees in a prospective group to fill out a health history questionnaire. As health insurance, we're assuming that the employees are completing these questionnaires with accuracy, thoroughness and honesty. We've observed, especially if you talk with medical underwriters who are reviewing these forms, that frequently the questionnaires are incomplete or missing key details. The medical underwriter commonly makes follow-up phone calls in an attempt to obtain the missing information. That step adds time and cost to the underwriting. For some unknown reason, speed of turnaround in making your decisions creates a competitive advantage. If you have a process that slows down the time you take to make your final rating decision compared to your competitor, it sets you at a disadvantage. It's almost as important to be quick as it is to be cheap. If you have a process where brokers and agents can get things turned around quickly, it helps you out. Most important, this process leads to omissions in information and sets up a process where underwriters are quoting rates that do not match the underlying health risk.

In early 2004, Reden & Anders worked with a national insurer to study the impact of incorporating pharmacy claims data into the small group underwriting process. It was a retrospective study of groups that the carrier had already underwritten and sold. We ran members through a pharmacy data tool and got information that the medical underwriter then compared to the information it had at the time it initially underwrote the group. Then we checked to see if incorporating this pharmacy data improved the results. We considered some simple and straightforward questions. Did the pharmacy data reveal undisclosed information? Did the pharmacy data validate the information the applicant had self-disclosed? Did the pharmacy data have a financial impact on the business? Is this data effective in assisting medical underwriters in matching premium to risk?

We put through as many groups as we could in the time period that we allotted for completing the study. We did 160 groups, totaling 2,281 members, from six different states. We broke down the groups for groups two to nine and 10 to 24. For this particular carrier, in many states once a group got to be 10 lives and above, they switched from what was referred to as a long-form questionnaire, which had a lot of detailed questions, to a short-form questionnaire. In a lot of markets, the competitive pressure won't allow you to do a complete, detailed, long-form questionnaire as the size of the group increases. One of the things we wanted to measure was whether there was a difference in adding the pharmacy data into the process for those groups where the employee questionnaire was a short form with less-detailed questions.

When you're using pharmacy claims data, you're going to be inputting information into the tool, including the member's name, gender, date of birth, zip code and, if you have it, Social Security number. I should mention, because the concern comes up, before you can enter anything into the tool, you have to have a release from the member giving you authorization to go after that information. There are strong Health Insurance Portability and Accountability Act of 1996 (HIPAA) concerns here. Without HIPAA-qualified language, we're not going to allow you to access the tool to gather information. We work with clients to make sure that the HIPAA-qualified language is incorporated into the questionnaire.

When you enter members into this pharmacy claims tool, you're going to get two types of hits. The first thing we do is verify eligibility. If a member happened to be covered by more than one health plan, it won't identify the health plan, but it will show you two different periods of coverage. One type of hit identifies that the member was, in fact, in the data from the eligibility but did not have any pharmacy prescriptions. The other type of hit will also confirm the member was in the data with eligibility and will give you a detailed listing of the prescriptions that that particular member was issued over the previous three to five years. From an underwriting perspective, both are valuable pieces of information.

I don't know how many of you have underwritten small group, but I can tell you that for the stereotypical seven-employee group where all seven questionnaires come back completely clean, where every question is answered no, most underwriters won't believe it. There's a tendency to think that members have gotten forgetful, or perhaps are less than forthright. I hope I don't offend anyone with this comment, but we've observed that the completeness of the answers on the questionnaire often varies a bit depending upon whether it was a man or a woman who completed the questionnaire. It's a stereotype, but men don't do as thorough a job. For whatever reason, in our culture, females tend to understand the health care concerns of their families more thoroughly than men do.

In any event, when you go into the data, one of the things that you're going to want to track is how many times when you enter a member into this database do you get data back? On this particular study, it was an average of 65 percent of the time, about two-thirds of the time. I was also interested in seeing whether there was a significant variation between employee, spouse and covered children. The response was relatively strong across all categories. Of the 2,281 members entered, 5 percent had a previously unknown condition that was identified as a result of our pharmacy claims data. Again, these were groups from whom we had already received a completed employee questionnaire. The employees were being forthright and honest and thorough in telling us about their health status. But for 5 percent of them, we got information from this tool that wasn't disclosed by the member. The most predominant illnesses and diseases that we picked up on were mental health and asthma. Initially I was hoping that we were going to pick up a bunch of hemophiliacs or people with growth hormones or some kind of big "aha." What we saw instead were incremental changes. But in an underwriting process in a competitive environment, those incremental changes are worth a lot of money. For 173 of these members, or 8 percent, our tool identified a drug they were taking that they didn't identify in the health questionnaire.

This particular carrier didn't use a debit model. It used a process where it would identify ongoing claims cost as a result of illness or disease that was self-disclosed. One of the things we investigated was the number of groups for which the medical underwriter would increase or project an ongoing claims amount enough to change rates as a result of the new information coming from those pharmacy claims. For 22 of the 160 groups, the information coming from the pharmacy claims tool was significant enough that the medical underwriter would have changed the rate tier. As you might have guessed, for groups where we were getting a short form or information that was less detailed, this tool provided information that if the underwriter had known it at the time it assessed the risk, it would have quoted a different rate for a little over a third of those groups. For groups with the long form, we would have changed rates on a little over 10 percent.

For 41 percent of the groups, we got new information that the medical underwriter changed its projection of ongoing medical cost. For 44 of those 66 groups, the change in cost was not large enough to have the underwriter change the rate tier that was quoted, or the rate tier that was quoted was already at the maximum. Having this additional information was nice to know, but since the underwriter was already at a maximum rate level, the rates could not be changed. Twenty-two of the groups would have received higher rates than were originally quoted. We would have rated up about a quarter of the groups of 10 to 24 and about 9 percent or 10 percent of the groups under 10.

If we looked across all 160 groups in the study at the additional premium that we would have collected as a result of the rate-up, it would have averaged 4.5 percent. The rate change just for the 22 groups where we would have wanted to change rates was 22 percent. When we took this additional premium, annualized it and divided it by the number of hits, the additional premium per hit in this study exceeded $100. In the industry today, if you go out and look at these tools, they're commonly priced on a per-hit basis. So one of the things we're looking at is whether a carrier achieves a strong return on investment. To keep it apples to apples, we compared a per-hit increase in premium to a typical per-hit cost.

These are 160 groups that this carrier had sold. One of the unintended outcomes of the study was that 11 percent of these groups provided new information that was disturbing enough to the medical underwriter that it's pursuing fraud. In its opinion, the member left out information that was so significant that it would have altered the way it priced the group. Because these were groups that sold, it was going to go back and see whether it could take this member or this group through its fraud procedures.

The use of these data did have a positive effect on underwriting. Underwriters could more accurately match quoted rates and project medical expenses, creating a strong ROI. I think this tool will have broad application to small group carriers. The tool initially had its genesis not on a group basis but an individual basis. Three or four years ago, the tool was used to help life insurance carriers underwrite high face value life insurance policies. Right now, probably about two-thirds of the usage within our tool is for companies that are underwriting individual products to assess anti-selection potential in the individual product underwriting process. That is a process where, again, attending physician's statements (APS) are frequently requested and phone calls have to be made. This creates a process that's not efficient and often doesn't give the underwriter complete information.

Adding pharmacy data into the individual medical underwriting process gives the underwriter enough information so that it can more accurately assess risk and decline the application. It can rider out illnesses or diseases. It can rate up. It allows the underwriter to be more informed as it's making its individual medical underwriting decisions. Adding the pharmacy claims data in the process will improve your financial results. You can identify those groups that are substandard and preferred risk. It helps to automate the renewal process, freeing up your underwriters' time so they're spending less time on the high cost claim analysis. All of this is going to lead to an improvement in your medical loss ratio.

When we've done ROI studies, the numbers are a lot softer in terms of the predictive modeling implications for renewal business. In general, we think you can improve your expense ratios or your loss ratios anywhere from 0.25 percent to 1 percent. But because our health-care dollars have gotten so high, a small improvement in loss ratio yields pretty significant total dollar savings.

Thank you very much. At this time, let's address some questions.

MR. LAURENCE R. WEISSBROT: I have not so much a question as an observation. This is something that I always fought about with the salespeople. Bad experience is more predictive than good experience. It's not a normal curve. If it's down toward zero, you can't rely on that, but if it's up, it's predictive.

MR. MINNICH: Unless it's acute and one time.

FROM THE FLOOR: If I bring this to my block of business or bring these tools to my underwriting, how would you characterize the long-term goals and the short-term goals that I should be looking at on this block? You say it'll improve your multiple location risk (MLR), but at the end of the day I'm going to price my block to a set MLR. Are you saying that I can get the same amount of business you talked about? The goal now is to grow business; that's what underwriting cycles are all about. But this is an attack at MLR. Are you saying that we can at higher profit margins grow business with these tools? What are the long-term and short-term goals?

MR. MINNICH: Great question. I would say the first goal is to allow your underwriting decisions to be as informed as possible. I'm a big believer that if your underwriters are informed, they're going to make the appropriate decision. As more and more of your competitors are introducing these tools into their process, their underwriters will be smarter, and your MLRs will suffer because you're going to see an erosion of your in-force block. Your better risk groups are going to be captured by your competitors at renewal, and you're going to underprice some of your substandard risk. All of this leads to a situation where it should free up additional surplus or margin. One of the questions you're going to have to answer is whether you are going to pump that additional margin into more aggressive pricing or let it flow to your bottom line.

One of the challenges that I've observed, which may be something that you need to be aware of, is that there seems to be a disconnect in a number of health plans where underwriting has a certain budget, and the cost for these tools is typically assigned to that underwriting budget, and yet the improvement in MLR flows through in another part of the organization. What I've seen sometimes as a stumbling block is that underwriting leadership can't get authorization to purchase these tools because they're not budgeted. You need to get somebody at a senior management level who appreciates the fact that you're going to be increasing your underwriting budget, but you're going to be reducing your medical loss ratio, so it's a positive for the health plan. You need to acquire it at a senior management level, because I've seen it happen more than once where underwriters or actuaries get enthused about it, and the roadblock is they haven't budgeted for it.

MR. WILLIAM SARNIAK: My question is related to the 6 percent of the people that one year represent over 50 percent of the cost and the following year are back to 24 percent. What's your recommendation in the renewal process when you have the time lag between the claims experience and the rating? We're pricing January 2006 rates using mainly 2004 data and a little bit 2005. Have you done studies on that or have any recommendations?

MR. MINNICH: Not surprisingly, I would recommend predictive modeling as a tool that will help you with that. If you're looking at selecting a predictive modeling vendor, make sure that this vendor has the capability to do a 12-month base period or gap and then your target period. Some of the tools that I've seen out there will do a 12-month base period to predict the next 12 months. As you've identified, in a renewal process that won't help you. Some of the vendors do build in that gap so those products do a more accurate prediction for a renewal process.

MS. RUTH ANN WOODLEY: In the work you did on new business using the pharmacy data, were there groups where you didn't believe that every member had a clean application, and so you overrated that group and lost them? Did you look at whether there were groups that, looking back, you would have rated lower than you did?

MR. MINNICH: It's something we wanted to do, but because this particular carrier wanted to do a retrospective study, the only groups that we input were groups that it had sold. But your point is an extremely valid one. Other clients that we've worked with have done a prospective or concurrent study where they're introducing this process over a 60- or 90-day basis. Certainly that is one of the observations, that underwriters feel more confident that that seven-employee group is a good risk and are willing to give it a more aggressive rate quote. At least in theory, you would think your quote ratios would improve on that particular segment, but our study did not address that.

MR. ROSS WINKELMAN: I think prescription histories are exciting and interesting for new business, but I'm a little skeptical on renewal. I think you're saying in the presentation that the new business methodologies for prescription histories are better than renewal methodologies. You made statements about the 65 percent hit rate on prescription histories and, as well, the fraud and the nondisclosure on new business applications. One thing we looked at is small group medical underwriting guideline debit point assignments. When we use actual claims data, the number of debit points is much, much higher than you get on new business applications, as much as five times higher. Is that what you're saying, that new business methodologies are better than renewal and more accurate?

MR. MINNICH: That's a good question, and I'm sorry if I was not clear. For new business, we're forced to underwrite without having access to claims data because when you're doing new business underwriting obviously somebody else has been processing the claims. Virtually nobody that I know of gives a group of less than 50 the historical claims data. So in the underwriting process, you're relying on information the member self-discloses. You wouldn't use the same pharmacy claims tool to do a renewal underwriting. You'd use a standard predictive modeling tool, and that tool is going to incorporate, the fee-for-service medical claims if you have access to it, as well as the pharmacy, as well as lab data. It's going to take advantage of all the information you have to come up with this predictive risk score. So it's two different tools that you'd use, one for new business, and one for renewal. For new business, as I mentioned, you're trying to fill the gap because you don't have access to the claims experience. At renewal you do have access. As I've worked with carriers, the predictive modeling tool at renewal is added to loss ratio experience in age/gender. Many carriers are looking at doing a three-tiered approach to credibility, not ignoring the claims experience but supplementing it with this new information from a predictive modeling tool.

MR. GEOFF SANDLER: I also wanted to ask about renewal business. Certainly new business is important, but from an insurer's financial results perspective, financial results are a lot more driven by renewals than they are by new business. I was wondering, given that, you do have some of the actual claim data at renewal that you can apply some of the same techniques, such as looking at prescription drug usage, how amenable do you think predictive modeling is to being applied to renewal blocks of community-rated business, for example, treating them as if they were one large group?

MR. MINNICH: Certainly in states where you can't do health status adjustments for small groups, applying the tool on a block basis will add value. You will be able to see whether the underlying risk of your block is improving or deteriorating, and whether it will allow you to do hopefully a better job of understanding that risk and making your pricing decisions. The last comment I want to make is that on the new business side, this tool currently is used just on small group, because you need a member authorization. The tool would add value in the above-50 groups, but one of the challenges that we're faced with today is most companies can't get an employee application once they get over a certain size threshold. On renewal, one of the improvements that we've seen over the past two or three years is that the tool will add value in a renewal process above 50. I primarily focused on small group, but if you get the tool, I would suggest incorporating it into your renewal process for groups of 50 to 200 or 50 to 300. Supplementing your current claims loss ratio approach will add a lot of value.

MS. WEYUKER: I'm going to talk about a slightly different angle on predictive modeling applications, focusing more on how to choose a good predictive model, given different kinds of applications. I'm going to go through some of the history of predictive models and how we got to where they are today, why you'd want to use a predictive model, testing the model, how to discern whether or not a model is a good predictive model and some applications of predictive modeling. I'm also going to spend some time talking about data. I feel that the impact of data and data quality on a predictive modeling application can't be overstated or overemphasized.

The people who have been around for a while may remember back to the 1960s when things were much simpler. This is before electronic data existed. Age and sex models and family status were the variables most commonly used by underwriters, along with geographic location and welfare status. There was also some information from survey data using self-reported health status. Back in the 1960s, it was used primarily in underwriting and pricing individual health insurance premiums. The models were good for the times, but with the access that we currently have to technology and electronic data, these models from the 1960s aren't good compared to what is available in the market now.

Let's fast-forward to more recent times. In 1984, which is a good demarcation starting point, research began on so-called health-based risk models. A lot of the interest was generated by the federal government, then called the Health Care Financing Administration (HCFA). In 1997, the Balanced Budget Act was passed. This piece of legislation mandated risk-adjusted payments for Medicare+Choice, implemented by Centers of Medicare and Medicaid Services (CMS) for the year 2000, based on a limited, unsophisticated model. If the federal government puts its stamp of approval on some type of method or technology, it's often adopted by the commercial world. I think that's part of what has happened with predictive modeling or risk adjustment. Since it's been blessed by the federal government, you can see that it's being absorbed by the commercial world.

There were other early adopters of risk adjustment. In the 1990s, an organization called Health Insurance Plan of California (HIPC)—it's now called PacAdvantage, which is a small group purchasing coalition—implemented a homegrown risk adjustment model that was created by what used to be Coopers & Lybrand. Since then, PacAdvantage changed to using a commercially available risk model, but it used to use a homegrown model. State of Washington employees used risk adjustment in the 1990s, as did the State of Colorado's Medicaid program. These were all for risk-adjusted payments.

Let's look at some other seminal events. In 1982, Medicare beneficiaries were allowed to enroll in HMOs. HCFA started funding a lot of serious research on risk adjustment in 1984, not just for the Medicare market but for the commercial market as well. Some of these risk adjusters were called things like diagnostic cost groups (DCGs), ambulatory care groups (ACGs), chronic illness and disability payment system (CDPS) and others. In 1993, the Clinton health care initiative was focusing on using risk adjustment in the non-Medicare populations.

In 1996 was the first SOA study that compared several different risk models and risk adjusters. This is an interesting study and is available from the SOA if you'd like to take a look. Risk adjustment was mandated for Medicare HMOs in 1997 and was implemented in 2000. The model at the time, considered to be sophisticated, was called the "principal inpatient diagnosis model." It used one diagnosis from one inpatient stay per year. It wasn't a sophisticated model, because it was using a small amount of medical data. It was mostly still an age/sex model. And, as some of you may know, this risk adjustment is being phased in, so that in 2000, 10 percent of the rate paid to Medicare HMOs was risk-adjusted. The other 90 percent was done the same way it had been done with age/sex and the other variables.

In 2002, the SOA put out its second study comparing several different risk adjusters. A comparison of the 2002 and the 1996 studies shows a big advancement in the area of risk adjusters and the models out there and their ability to predict risk. In 2004, CMS selected a more modern model, the so-called "all-encounter model," which uses diagnoses from inpatient and outpatient sites of care. Again, CMS is phasing this in gradually.

I'm going to use a slightly different angle to talk about why you would use risk adjustment or risk assessment. Predictive modeling can be a synonym for risk assessment and risk adjustment, so I use these terms interchangeably. I've had this discussion with a few actuaries, comparing risk adjustment as a predictor versus some of the other methods that actuaries have been using for many years. When comparing risk-adjusted predictive power versus age/sex versus prior claims methods, in businesses from jumbo groups all the way down to small group employer down to two employees, the risk-adjusted predictor beat out the other two methods in every single case except for employers of size two to 10. In that case, the prior claims method won out.

I'm doing another plug for the data. I think that part of the reason why risk adjustment and predictive modeling are becoming such hot topics right now is that the electronic medical data that are available, which is what is going on behind the scenes to make these models possible, have dramatically improved in quality over the past five years. You have to have good data as the basis for a good model.

I'd like to use some of the statistics to compare these three different types of predictors. In case some of you don't remember all of your statistics, an R-squared measure is one of the measures typically used to demonstrate that a risk adjuster is a good risk adjuster or if it's just an adequate risk adjuster. An R-squared measure measures on an individual basis unless you're using an aggregate R-squared, but most vendors are not quoting an aggregate R-squared. Starting at the bottom of the scale, age/sex has a pretty weak R-squared. It's usually less than 1 percent. If you're lucky, you'll get it at 1 percent. This shows that you're predicting 1 percent, so it's not a good predictor on an individual person. Prior claims methods show an R-squared of 4 percent to 5 percent in some studies that I've seen. Again, it's not great. Health-based risk models have demonstrated a wide range of R-squareds, from 8 percent to 50 percent. Some vendors apply different kinds of tricks to overstate the R-squareds a little. For example, you can truncate the claims at $25,000, or you can truncate the claims at $10,000. But the point is that the health-based risk models are superior predictors, at least using an R-squared measure. Another reason to use health-based risk models can be demonstrated in another metric called a predictive ratio.

Using a predictive model also allows you to even the playing field. If you remove risk from the equation, you can look at comparing health plans. For example, there are some health plans whose main expertise is cherry-picking. If one carrier that has a huge part of the market share is cherry-picking, that leaves other health plans to take up the rest, which could be a sick population, which is difficult to do on a financially viable basis. Another application is allowing providers to be paid on the basis of disease burden. Again, if providers are treating a population that is at risk and that has a disproportionate share of, for example, chronic diseases, risk adjustment allows those providers to be paid in a way that measures the additional disease burden that they are seeing in their offices or in the hospital.

Weyuker Slide 15 shows some predictive ratios. It is a good illustration of what an age/sex model looks like for different cohorts. On the bottom axis is the predicted annual cost grouped in bands. The far left-hand band is $250, and the far right-hand band is $40,000 or more. The vertical axis shows actual cost. To do this graph you have to have two sequential years of data. One year of data is the predicted cost, and the other year is the actual cost that occurred the following year. The line representing the age/sex model follows pretty closely to the bottom axis. The actual cost line is a curve. An ultimate predictor would be following that curve exactly. The age/sex model on the bottom does a decent job of predicting actual cost up until about $4,000 a year, and then it starts to veer off. So the age/sex model is not doing a great job at predicting the parts of the population that are expensive. The medical-based model follows the actual cost pretty well, and it certainly is doing a much better job of predicting than the age/sex model.

Some of these statistics may not be what you would see in marketing materials from vendors. The SOA 2002 study showed results of three predictive models. The DCG model showed an R-squared of approximately 15 percent. CDPS similarly showed about 15 percent, and ERGs also showed 15 percent. There is not a meaningful difference in predictive power of these models. They are all doing a pretty decent job, especially when compared to age/sex models or prior claims methods. Again, you can see these R-squareds at a much higher level depending on how the model is tweaked and if the claims are truncated and so on. Also, this study is from 2002. Some new versions of the models have come out since then, so I'm sure the R-squareds have improved from the results stated in the SOA study.

Weyeuker Slide 17 shows an example of the predictive ratios. Predictive ratio is predicted cost divided by actual cost. Some people show it the other way around, but here we're showing predicted cost divided by actual cost. A predictive ratio of 1.0 means that you've predicted actual cost exactly. This slide shows three chronic conditions and how some of the different predictors perform using a predictive ratio metric. For the first condition, depression, the first bar represents the age/sex model, which gets a predictive ratio of less than half. That model is not doing a great job of identifying those people who have a diagnosis of depression. The next bar reaches 0.58 using an inpatient-based diagnosis model. And the next bar, which reaches 0.96, is performing the best in all three examples. This is the diagnosis model using data from all sites of care. It's using diagnoses only. And 0.96 is about as close as you would get to an exact prediction. The next bar is a drug-based model, which gets 0.79, and the following bar, a drug plus inpatient diagnosis model, achieves a ratio of 0.82.

For the two other chronic diseases, diabetes with complications and asthma, the models are performing similarly. The all-encounter diagnosis model is doing the best job of predicting cost. The drug and the drug plus inpatient models are doing a pretty decent job too. This is important, because sometimes people like to choose the drug-based model because it's easier to get the data. There's been a lot of focus by some employer groups to purchase drug-based models since the data are so much easier to obtain. You also don't have to wait a year or six months to get claims; there is less of a claims lag with drugs. Consequently, drug-based models have a lot of appeal, and this shows that they do a good job. The highest number you would get theoretically is a predictive ratio of 1.0. But you can over-predict. I have seen a predictive ratio greater than 1.0, but it didn't happen with the data in this case.

For the data flow in using a predictive model, it's probably safe to say that the input for all models includes age/sex information. If you're using a diagnosis-based model, the input will also require diagnosis information. If you're using a drug-based model, you also have to input drug data. Typically the models are designed to require one year's worth of both enrollment and claims data. Some models require other variables to be input, including the frequency of claims and the timing of claims. There's also some use of lab data, as well as other variables including the total cost of health care consumption for each person in a given year. The output typically is clinical categories if using a medical model, or drug categories if it's a drug-based model, and the ultimate output is the risk scores for each person.

As far as data implications are concerned, as Jim already mentioned, HIPAA is crucial. If you work for a carrier and release the data without getting the appropriate consent, there are stiff fines, so it's important to abide by the HIPAA laws. In addition, even though HIPAA is a federal law, some state laws can be more stringent than HIPAA. I'm from California, so I know more about California's confidentiality laws. California's laws, which have been on the books for some time, are more stringent than HIPAA. For example, California law is strict with respect to mental health data. You're not allowed to use mental health data unless you get additional sign-off, and there are other constraints with respect to abortions for minors and so on. I highly recommend contacting a HIPAA attorney if you have questions about this.

There are a few things that affect data quality. One thing that I still notice is that self-funded plans don't have the data quality. A self-funded plan often will collect data on someone if that person had an encounter with a medical system. But frequently if an enrollee has not had any health-care consumption in a 12-month period, he's not in the data. This is a problem because the age/sex adjustment has to be calculated on everyone in the population even if you had no health care, because the fact that you had no health-care consumption is part of the data. It shows there's a healthy person in the population. So it's important to make sure you have everyone in the enrollment data if you're going to do risk adjustment.

In addition, data bias and dirty data can have significant impact on any analysis that you do with predictive modeling. Let's say you're comparing two different medical groups. Medical Group A only keeps one diagnosis per claim in its computer systems. It may be in a rural area and not have great computer systems. But Medical Group B is in downtown Los Angeles. It's sophisticated and keeps up to five diagnoses per claim. If you did an analysis on this, Medical Group B would look sicker than Medical Group A by a disproportionate amount. You would get an unrealistic view, just because of computer systems. This could lead you to an incorrect conclusion. Similarly with hospitals, a hospital in a rural area may not keep as many diagnoses as a hospital in an urban area.

So if you're doing a comparison, it's important to make sure that the systems are keeping a similar number of diagnoses. If they're not, you need to intelligently truncate the number of diagnoses you're going to analyze. In a case where we used the so-called all-encounter model, which uses diagnoses from all sites of care, when we kept only the primary diagnosis, we got a risk score of 0.79. When we kept the first two diagnoses, we got a 0.86, which was a change of 9 percent. Going on three diagnoses gave us 0.92, four diagnoses gave us 0.96, and five diagnoses gave us 0.98. The overall change in risk using all five diagnoses versus keeping just one is a change of 24 percent. If you were doing a product where you said the overall risk score of this population is 0.98, that's different from saying it's 0.79. It could lead you to some different conclusions about the risk of a population.

In a similar scenario, let's say we have two hospitals. Again, Hospital A keeps, at most, three diagnoses per claim, and Hospital B keeps, at most, five diagnoses per claim. Hospital A has a risk score of 0.95, and Hospital B has a risk score of 0.92. Based on the impact of missing diagnoses, let's say that Hospital A kept all five. We could apply some math and assume that the risk score might be 1.012. Now, we don't know that, but this is why it's important to have all the diagnoses.

Some strange things that can happen with drug data that affect its quality. If you decide to use a drug-based model, having clean data is important. You can end up with some odd incorrect conclusions if your data are not clean. For those of you who may not look at drug data a lot, in an 11-byte national drug code (NDC), the first five bytes represent the manufacturer code. The next four bytes represent the chemical itself, and the last two bytes represent the dosage. The NDC 00002 is the manufacturer Pfizer, and let's say the chemical code is 0445, and dosage is 01. By the way, if you tried to go back and match this, the 445 is meant to be an example. I don't remember if that's the actual number. It may not be Acyclovir, but this is meant as an example. So let's say this is Acyclovir at a dosage of 400 milligrams twice a day.

I was working with a data analyst who was trying to save computer space, and so he decided to delete all leading zeroes. You can see something strange that could happen as a result of the analysis. Since he was keeping all three of these fields separately, he decided to truncate the Pfizer manufacturer code of 00002 to just 2. Similarly, he got rid of leading zeroes in the chemical field and in the dosage field. If you reassemble these three into the NDC, you can end up with different implications about what this drug is. If you're running a drug-based model, the first nine bytes of the NDC are necessary to run the model. The model could think that this is Acyclovir, which is what it is, or the model might assume the missing zeroes are at the back, which would turn out to be Minoxidil. This creates different implications. Minoxidil, just an over-the-counter thing, does not imply a person who's ill, and Acyclovir could imply someone with AIDS. There are different disease burden implications with these two different results from messy drug data. This shows how important it is to have clean data and data that are free of bias so that you can come to conclusions that are as correct as possible. I can't overstate how important it is to be intelligent about your use of the data.

The applications of risk adjustment are limited only to the creativity of an actuary. Risk adjustment has been used for provider profiling for at least several years, as well as in underwriting and pricing and provider reimbursement. It has been successfully used in resource allocation. Predictive modeling works well together with disease management programs. Another area that doesn't get a lot of attention is using predictive modeling in incurred but not reported claims (IBNR). Predictive modeling is another way to increase the specificity of IBNR factors and maybe do a slightly better job of getting your reserves right. It is also used in trend analysis.

In conclusion, risk adjustment, when properly applied—and I'm big on that—is superior to many techniques that have been used, especially in the past in the health sector. Improperly applied, or applied to dirty data or messy data, you can come to some incorrect conclusions. To avoid that, I recommend that you become knowledgeable about risk adjustment and how it works. I believe risk adjustment models are still in their infancy, and they will improve markedly as data quality improves. We may get R-squareds of 70 percent by adding artificial intelligence or who whatever improvements that may be on the horizon. This is being adopted in a big way by the health insurance industry just now, but it still has not been applied across the entire health sector. Many physician groups still are not using this. Many large employers are still not using it. Insurers, HMOs and even parts of the government sector are still not yet using predictive modeling.

MR. THOMAS MESSER: I have two questions. First, in underwriting, typically we might use 2004 data in 2005 to predict 2006. What happens to the predictability if you spread it over two years rather than just one?

MS. WEYUKER: This is something that some predictive modeling companies have been focusing on. The predictive power is much less in the second year than it is in the first year. The models primarily have been developed to have a much higher predictive power in year one than in year two. So there's still a lot of improvement to be made in doing a two-year prediction.

MR. MESSER: My second question is, in terms of aggregating the data, if I have one person, it looks like it's a pretty good predictor. If I have 100 people, is it better than age/sex? If I have 1,000 people, is it better than age/sex?

MS. WEYUKER: That's a great question. If you have a million lives, an age/sex model is a great predictor. It depends on how many lives you have. From what I've seen, if you have 2,000 lives, predictive models are still superior to age/sex. You have to have a very large dataset for age/sex to be just as good a model.

MR. JOHN F. FRITZ: I'd like to tie your two areas together in terms of the privacy issue and the HIPAA implications that you mentioned, Lori, and, Jim, how you gather the data? How do you get the pharmacy data, for example, and continue to get it in light of the HIPAA regulations and keep that current? If we have to take names and all that off the records, how do you then use that in the underwriting? Are the data getting stale because you can't keep refreshing them?

MS. WEYUKER: If you deidentify your data, that makes the data usable, as well as, as Jim mentioned, getting the insureds to sign off on the waiver. If you deidentify the data and aggregate it, you can certainly do this, and it's not a problem. The HIPAA laws state some of the things you have to do. You have to use three-digit zip codes. Obviously you can't have names, Social Security numbers, age or sex. You can't have things that obviously identify the person. Jim, do you want to add to that?

MR. MINNICH: For the new business underwriting, clearly the data providers to the companies that are selling this are concerned about HIPAA. They won't allow you to enter a member into their data unless you get that HIPAA release. When you get lawyers involved, they disagree about what appropriate HIPAA release language is. We've walked away from a couple of deals because we didn't feel the client's HIPAA was strong enough. On the new business side, in the employee application where the employee is filling out a health questionnaire, we'll build the language right into that questionnaire. The employee will authorize us to go and ask the pharmacy for that data, which they have the right to. HIPAA gives all of us the right to get our information; the employees are just giving us access to that.

MS. WEYUKER: In addition, in California sometimes the release has to be signed by the dependents as well. For example, you have a dependent who's a minor who's having an abortion. That minor has to sign the release herself. Similarly with mental health claims, if the dependents are having the mental health claims, they must sign the release themselves.

FROM THE FLOOR: This is similar to the previous question in terms of the lag. When all these studies are done, are they using data right up to the date of the study? If you have an underwriting lag of four to six months, how fast does your predictive power deteriorate? You said the second year predictive power was worse, but, as you move through, is there any feel for that?

MR. MINNICH: As Lori said, you get your strongest predictive power when it's consecutive—a 12-month base period followed immediately by a 12-month target. I've done several validation studies where we built in a three-month gap, and we've seen the predictive power stay pretty strong. It does drop a point or two. It is important as you're looking at vendors, as I mentioned earlier, to make sure the vendors have the capability to put that gap in there if you're going to be using it for underwriting.

MR. ROGER D. LOOMIS: My question is for both of you, about calibrating the risk adjusters specifically for small group underwriting. In a small group-underwriting scenario, there's a relatively narrow rating band. What you're most interested in is figuring out where within, say, plus or minus 35 percent a particular group should go. However, when you're calibrating these things it seems to be driven by the points that are extreme, way outside of that band if you're using ordinary least-squared regressions. How do you go about calibrating it so it's based upon the data points that are relevant to the rating band?

MS. WEYUKER: In my opinion, when calibrating a model most of the models that are available by the commercially available vendors are already calibrated for a specific market. They're either calibrated for a Medicare market, for a Medicaid market or for a commercial market. You have to have a huge volume of data to recalibrate the model in a way so that the coefficients are stable. I would suggest to you that you need access to a large database that's just small group if that's the way you want to do it.

MR. MINNICH: The output, as you suggest, will in most models give you a range of 0.5 to 50 or 50+, which are significantly outside of your plus or minus 35. So the next step that needs to happen, either internally or hiring an educated consultant to come in, is creating a table that would suggest if you're in a plus or minus 35 and you have 18 tiers of rates going from 0.65 to 1.35, a predictive risk score under 0.5 will get you the first table. Then 0.5 to 0.7 will get you the second table, et cetera. You can't just take the output and immediately multiply that by your rates, because in many states the small group reform won't allow that kind of flexibility.