RECORD, Volume 29, No. 2*

Spring Meeting, Vancouver, B.C.
June 23–25, 2003

Session 103PD
Predictive Modeling—Current Practices and Future Applications

Track: Health

Moderator: DAVID AXENE
Panelists: THERESA KEANE-NORTH†
RICHARD LILIEDAHL‡

Predictive modeling continues to show great promise as new applications are uncovered and approaches are refined. A panel of actuaries and clinical consultants provides an assessment of current and future applications of predictive modeling techniques. Attendees gain an understanding of current practices and future applications of predictive modeling.

MR. DAVID AXENE: Our first speaker is Theresa Keane-North. She's the Director of Performance Reporting. She's been in group health for five years, and she's responsible for the product line performance reporting, underwriting data and risk adjustment implementation. I don't know how many of you are aware of Group Health, but it's one of the oldest HMOs in the country. It started off as a stock model but is now sort of a mixed model. They have a variety of types of providers, either employees of the plan versus contractor providers, and they have been an industry leader in a variety of areas. We're pleased to have Theresa with us to tell us what they're actually doing. Now, we will quickly gravitate to the more theoretical side of it, but we thought we would start off with a very practical presentation about what people are really doing.

MS. THERESA KEANE-NORTH: We at Group Health have been interested in risk adjustment for a very long time. One of the reasons for that is that the old traditional business model of let's figure out how to have the skinniest benefit plan

---

†Ms. Keane-North, not a member of the sponsoring organizations, is Director of Performance Reporting at Group Health in Seattle, Wash.
‡Dr. Liliedahl, not a member of the sponsoring organizations, is Chief Medical Officer at Axene Health Partners in Temecula, Calif.

and the skinniest network to attract the best population is really not our style, so we actually are very excited about the introduction of the new technology in risk adjustment and the ability for us to begin to see employer groups and other external payers use risk adjustment to compensate us.

What I'm going to talk about primarily today is the use of risk adjusters in the commercially insured populations. We're doing many different things at Group Health regarding this. We are exploring the idea right now of using risk adjustment in our underwriting model. We are in the process of redoing our model now; we started playing with the idea a little while ago about actually using risk adjustment as a factor in that model and decided that we're not quite there yet, but we're getting pretty close. One of the reasons we wanted to do this is that it's certainly better and has better predictive power than age, sex or community rating, and it's more forward looking than the traditional experience rating. It does a much better job of eliminating one-time events, so the preemie baby or the car wreck—those things get excluded in our risk adjustment calculation, whereas in traditional experience rating as you're pricing a group, those dollars are still in there.

Also, it can give us a little bit more of a forward-looking view of what's going to happen with some of the chronic conditions. So in a commercially insured group where your experience rating period is from January to December and somebody's diagnosed with a cancer in November, you're not going to know in a traditional experience rating method that that expense is going to hit until the following year, whereas in a predictive risk adjustor you would actually see that earlier on.

We have taken the position at Group Health that we're not really excited about the idea of using risk adjustment in small groups under 50 to price specifically. We do like the idea of there being some sort of a large pool where maybe carriers are risk adjusted, but currently in our states we actually do not have the flexibility to use any sort of health status–based underwriting factor in the small group area. However, we are using this to evaluate our small group pool. We've done some studies by benefit plan, by geography, looking at our chambers of commerce and associations as well as looking at the churn in our small group pool. What's been very interesting is some of the things that you would expect to see with the benefit plan, for example, higher-risk people choosing the higher-benefit plan and the lower risk with the lower-benefit plan: those things proved out to be true. What was really nice about this was for the first time we were able to sit down with our product development people and say, here's the difference in health status as the result of your benefits. So those who had a rich pharmacy benefit, we had people who had a lot of chronic conditions that required drugs.

Also, another piece that was very interesting to us is our work on the chambers and associations. It was very clear to us after taking a look at the difference in management of these chambers and associations. Those that were well managed by a credible broker who actually paid attention to what was going on in the pool performed much better than those that were not managed well and that were just

thrown together as a group of people trying to get insurance. Again, we feel strongly that we shouldn't be using it to underwrite small groups.

Now, mid- and large groups, however—we feel the risk adjustment is excellent in that setting. We did a study last year, and we're actually going to be doing the same study again this summer, where we evaluated 83 groups, and out of those 83 groups, 55 came closer to the actual expense in the following year using a risk adjustment methodology rather than the experience rating methodology. So our director and VP of actuarial underwriting really wanted to see a more certain view, but we've made some improvements to our data since then. We had a gap in our claims system, and we weren't getting all of the correct diagnosis captured. That has since been resolved, so we think we're going to have a much better outcome of this study this summer as we do it.

As I mentioned, we did some comparing of scores by benefit plan, which was quite interesting. We also looked at some things by geographic areas. Where we were a network provider, we actually had higher risk because they could choose the same providers within that network and still have a high benefit; however, we were lower risk in our own network for certain segments of population. However, for the Medicare population we were actually much higher risk because people tend to stay with their family doctor. So where we use a staff model, the risk was generally lower for those populations that tend to churn a lot, but those populations that stay with us, they stay over time because they just like to hang out.

We also did some benefit comparisons to competitors and found that we are very rich in benefits as are many HMOs, and with that we made some decisions to start offering some deductible and co-insurance plans within our HMO. And then I already mentioned the churn analysis work. Even though we decided not to actually use this as a factor in our model, we made the decision that we wanted to at least have the information available to our underwriters as to what the risk profile was of the group, and what we were also able to do was give them kind of a heads up on, are your number of diabetics increasing in this population, are your number of people who have heart conditions increasing? Is it that you've got a situation where you're just having a lot of the kind of the exams and routine stuff that's going on, and is that driven by benefit? There are lots of different questions that we're able to ask and answer using these tools.

We're also using this information for rate negotiation. Very often we go out with an increase, we may have an actuary or two look at it or a consultant and say, you know, that's just a little high, and with this we're able to sit down and talk about why. You've got more folks with cancer this year than you had last year, or conversely when we're able to actually give a below trend increase, we can say, you know what, your illness burden in your population has gone down.

Another thing that we've been able to do with one very large employer group in Seattle is sit down with them and talk with them about what is going on within their

group that is causing the high utilization or high use, high expense. This particular group had a very rich benefit plan, and if an employee could manage to get a note from the doctor, they could have the day off, so guess what, we had a lot of doctor's visits, especially on sunny days, by the way. So we were able to talk with them about that and have them have that as a tool as they go to negotiate with their union.

Also, one of the things we noticed in that particular group was we had really high incidence rates of diabetics, folks with anemia and other nutritional disorders. So what we did was we sat down and talked with them and said, so what are you serving in your cafeteria, and interestingly enough it was burgers and fries. So we started talking more with the employer group about what are some of the things that you can do to help people to educate people around their diet and nutrition. Then we were also able to find some situations where people were having cancers and a lot of lung-related diseases, and we may have been able to track those back to their working environment so that some other protections could be put in place for that.

There's a lot of external risk adjustment. We call it external risk adjustment, but it's going on, and I'm actually personally very thankful for this. We currently submit data to Medicare, Medicaid, Washington state public employees and Boeing, as well as a few other smaller groups now, where we're submitting diagnosis data, and they're actually using these data to risk adjust our premiums. That's been a real win for us because we had long believed that we had the higher risk in many of those populations, and it turns out in many cases to be true. So in the old world when you put down a filing, they'd get frustrated with us because our rates were higher than some of the other clients that were in there; we were able to stay on a risk-adjusted basis. Actually we looked pretty decent, so that's helped us tremendously.

One of the other things that we're doing at Group Health with risk adjustment is panel size adjustment; actually we worked very closely with DXCG. I know Marilyn Kramer's in the room here somewhere. Marilyn's the president of DXCG, and we actually pretty exclusively use their product for risk adjustment, and they helped us through creating a panel size adjustment for our staff model. In the old world we had just an age-sex adjustor, so physician panels were adjusted based on that age–sex methodology. It turns out that when you add illness burden to that, you get a much more fair distribution of actual work load. We have some doctors in our system who were saying, "We're really overworked, and I don't know how I get through my day every day," and as soon as we started taking a look at that, we found out they were right.

Other things that we do: product line analysis. Dave mentioned a little bit of the work that we did around small groups, but we actually are taking this a step further and looking at all of our products in our array and trying to figure out what we attract or who we attract when we offer richer benefits or when we offer larger networks or point-of-service (POS) products, and what would it look like if we did a

PPO and those kinds of things.

We're also in the process right now of recalibrating our model for selective measures. The DXCG model has 184 condition categories. How many here are familiar with that model? Not too many. It's a hierarchical and additive model. In the diabetes category t here are six subcategories, and somebody who has diabetes is listed only in the one that is the highest within that category, so it's hierarchical in that respect, but if they have a heart condition and they have diabetes, those two conditions get added together. What we're doing is we're developing weights for the number of hospital days, for example. So we'll have, in each diabetes category, what are the days for 10,000 members per year that you would expect to see for people with that disease.

We're doing a similar thing with emergency room usage. We're also going to be doing this with lab tests and consultant visits. So we expect to have a really rich physician profile at the end of this where we can say here's what's happening within your population, the number of days that you're producing, and here's what we think it should have been. Are you performing better or worse than others? Then we're also able to run some scenarios with this model, looking at the data that we're submitting to our external risk adjustors to make sure that we are being fairly represented within those models. So we make sure that it's complete.

MR. CHRIS SIKES: You made the statement that you would not recommend using this for small employers under 50, and I realize part of that is the regulatory environment, but for those 30, 40 states that do allow tiering, would there not be a reason to use it for small employers?

MS. KEANE-NORTH: That kind of boils down to at what point is it right to do this, and I don't know what the right size is for you to start using health status to risk adjust a group. From a philosophical standpoint, you could really prevent a lot of risk if you were able to use kind of a health-based risk assessment and underwrite specifically. It's kind of ethically what do you think is right.

FROM THE FLOOR: You've used it phenomenally as far as the extensiveness in many different areas. My question is: how long, how much time? You mentioned for many years, but I assume that the DXCG model has been fundamental in recent years, and so for how long and how much staff time?

MS. KEANE-NORTH: Group Health has been involved in risk adjustment, I believe, since the early '90s, and the risk adjustment that was going on prior to our purchase of the DXCG model was a pharmacy-based model developed by our center for health studies; Carl Fishman developed that model. One of the reasons we made the decision as an organization not to use that model exclusively is because many of our Medicare beneficiaries do not have a pharmacy benefit. So in order actually to do some fair risk adjustment for physicians, either we needed to cough up money for a pharmacy benefit, or we needed to change models, and we

decided to change models. We've had DXCG now for about two and a half years, and it's performed quite well. At first, I think it took us a week and a half, something like that, to bring the product up. It's pretty quick implementation, but, of course, we noticed we had some data issues, and we took the last year to resolve those; then we felt like 2002 was a really good baseline year for us. Our data are in good shape, working the way we want it to. We're in the process of recalibrations as I mentioned earlier; I feel like 2002's been a great year for us. In terms of staff, it's three others and me, and I'm working only part time on this. We have a population of about 600,000 beneficiaries.

MR. PETE REILLY: I'm particularly interested in your underwriting applications, and I had a question about data cutoff, the impact on the ability to predict and bring it into an underwriting process, specifically around how recent is the data you can actually use, the timeframe between when you get that and developing the indicators, and then actually bringing it into the actual renewal process. I understand there are some challenges around that, and I'd like to understand how you are dealing with that.

MS. KEANE-NORTH: Well, most definitely there are challenges around that. We made a decision that what we were going to run this model quarterly and produce scores quarterly. And we have three months lag time, so at the beginning of April we run for the calendar year, and then we run every three months. Now, does three months account for all of the lag that exists in our system? Probably not, but I think are we touching most of it, yes. Within the models, the way they work is an individual only has to be diagnosed with a condition one time per year. We've done some studies and looked at what happens. How much does it change the score if we add six months lag? It's really very minor. So we're waiting three months, and then we're running.

MR. ROBERT WARREN: Before I even ask my question, I counter, I guess, your objection to using it for small groups. Right now we're using loss ratio anyway, so we're using something and that might be better. My question has to do with your saying that this is a better predictor than age-sex, and I'd like to know how your studies have shown how much better and to what size groups. Also, I've heard contentions from people who are proponents—that they find that these risk adjustors are even better than experience after some period of time or before that period of time of group size, and I wonder if you have studied that as well?

MS. KEANE-NORTH: We actually have done studies, and I want to first say that the R square, which is commonly used to measure models and the scores for age-sex, I think, is right around 0.02, and on the concurrent models we're at 0.38, something like that, and for the prospective models I think we're up to 0.14. We're higher than that now? We're in the 20s. So there's a lot of work that's gone on to prove that the diagnosis information actually does add to the value. Of course, any time you add more variables to a model, assuming that those variables are credible, you're going to get a better outcome, so we believe that. In terms of

what group sizes and where is it credible, I personally find that as the group gets smaller, the risk adjustor is more valuable.

FROM THE FLOOR: It might be interesting for folks to hear because you and I are not actuaries what objections were heard and what was the learning curve to the actuaries in your plan and the actuarial consulting community in rolling it out? This is a health service and research model being brought to the actuarial community. How you might learn about the education process?

MS. KEANE-NORTH: We actually only employ one actuary at this point. We're hiring our second one now. His name is Brian O'Sheals; he is a pretty amazing person, and he really embraced the idea of risk adjustment and is really very excited about moving forward on that basis. We provide data to him on a regular basis for his analysis and his studies, and I don't think he's doing any studies any more now that don't include some sort of a health-based risk adjustment factor in that process. In terms of getting the rest of the underwriting staff in our organization to embrace it, I think that one of the things that helped a lot was our external risk adjustment. The fact is that by the year 2007, well over 40 percent of our revenue as an organization is going to be risk adjusted. It gives people some idea about maybe this stuff really does work, other people are using it, let's take a look at what this means to us.

MR. AXENE: Richard Liliedahl is a physician who spends a lot of his time helping people understand predictive modeling, particularly from a medical management or utilization management perspective, but Dick is one of those doctors who likes to play with actuaries. Dick was M&R chief medical officer for the care guidelines division, and he was also a consultant with Ernst & Young and now is off trying to do something different.

DR. LILIEDAHL: I want just spend about 20 minutes talking about predictive modeling from the clinical perspective. I know most of you may not be interested in that, but for those of you in organizations where you work with physicians, I may be able to shed some light on some of the things they're doing, and I'd also like to encourage you to think about those organizations working more closely together with the clinical people when you use these predictive modeling tools.

I'm going to talk a bit about the vendors that sell these products in this country. It's hard to go in depth, but for those of you who have purchased these products or are looking at them, you know it's a very complex issue trying to understand what the products are, what they do, what you get from them, if can you use them for more than one purpose, what these people are really talking about who use these terms that aren't actuarial in nature. I'm going to talk a bit about that. I'm going to briefly go over some current clinical applications. I'm going to talk about some new medical management applications, and then I'm going to talk a bit about some of the things I think they can be used for.

Now, one of the things I had an opportunity to do earlier this year was to take a look at a lot of the vendors in this country and the products they put out. Now, when we did this, we didn't look just at the actuarial perspective, we were also looking at the clinical perspective, and most of you are probably familiar with what you want to call risk groupers or risk adjustors, and so that's your primary knowledge about this. But we were trying to think about it in a bigger picture, because some of these vendors have products that overlap, and it's really hard to tell what they are. So we tried to create our own classification, and it's basically risk status, health status and treatment status.

The risk status is basically a classification similar to a risk or the things that you're most familiar with, and that's the probability for a future clinical or a financial outcome. That's the type of thing that Theresa talked mostly about. There's also health status, which sort of addresses the current health status of an individual and/or a group and their current need for clinical or financial resources. Then we have another classification called a treatment status, and this is a primary tool that's used by disease management companies who are in programs where they're trying to figure out if patients in that population based on data are getting treated according to current best evidence or guidelines out there. So, for example, someone with diabetes: you can look at the data and figure out if they're on the right treatment, having laboratory testing at the right time and so forth. But those are really three different classifications. These classifications are in this document that's on the SOA Web site. I guess what I'm trying to tell you is there are a lot of products out there that have many different uses, and if you go to select a vendor, know what you're getting.

I want to talk a little bit about the current medical management applications. Most of you are familiar with these, but as we looked around to see what was going on, I think the most common current application of these tools—and they're being used for many reasons in many applications and in many different ways—but the most common use is just looking at practice pattern variation, looking at a diagnosis and trying to figure out what kind of variation is going on there, and then trying to figure out what to do about it. So if you look at the way people are using these tools within medical management divisions, I think that's the commonest application. They're currently great tools to identify practice variation.

Disease management companies and those who use disease management vendors—their primary application is to identify at-risk patients to put in the disease management program. Some people are using them to identify high-risk patients who are yet not in a case management program. That's probably pretty obvious, but that means you look at what happened with the population in the last year as far as the risk or in those individuals who have high risk and aren't in a case management program in your organization. You identify those individuals and try to intervene in their lives and improve their health care costs and their health care for the next year.

From my viewpoint, when you look at these tools and the way they're used, in most of the organizations we've looked at, they have been around or heard present, etc., and by talking to the vendors, our belief is that in most of these organizations the medical management departments are really ahead of the underwriting side. Now, there are some exceptions. We work with somebody in the Northwest who happens to be in this room, and in that organization I would say that the actuarial side is ahead of the medical management or clinical side, but in general I think the medical management side is ahead of the financial side.

I just want to give you one current old application. It's being used currently, and it's a really old idea, but with these new tools that are available I think it's quite exciting that you can look at practice variation in a very different way. I want to talk to you about one case study to try to explain how people are using us, and this was a case where they used an episode tool to look at cost by particular episodes. There are many different vendors who have products who group things by episodes. The first thing they did was identify the physicians who spent the most dollars per episode per diagnosis, and in this specific case study, they are just simply looking at otitis media. Then once they identified all those high-cost physicians, they looked at the high-dollar categories where the money was being spent, and with these tools, you can look at lab, you can look at X-ray, you can look at pharmacy, you can look at surgery; and in this particular situation, their biggest dollars and their biggest variance was in the pharmaceutical area. So they simply identified the drugs on which a lot of money was being spent, and they instituted an educational program, both with patients and physicians to change their physician prescribing habits for certain antibiotics.

We did these things 16–17 years ago, and I worked at Group Health where Theresa does, so this is not new. What is new about it in my mind is that these tools identify like groups of patients, and that's really important to physicians, because when you work with physicians, they all tell you they have the sickest patients. Well, these tools that group patients by episode, I think, really makes clinical sense to physicians, and I'm personally excited about these new tools, having done this for 15, 16, 17 years, because I think it's going to have more impact on the use and variation as far as the way physicians treat patients and the cost. When you explain this to physicians and these episodes that group patients, it makes clinical sense to physicians how these patients are grouped. And if you have a good building block underneath all this, I think it's going to have more impact on the physicians, but it's also a better building block to create any kind of risk or a risk grouping.

Let's talk about another current application, and this is probably newer. This has been being done only in the last couple of years. People who are doing this are in the first phase of it or have only done it for one year, so this is another case study. The last example was from a health plan in the western half of the United States. This is from a health plan in the central United States, but what they're really trying to do is find some at-risk patients in their current population who weren't currently

under disease or case management programs. So they used the risk group and identified patients with high-risk scores, and they chose at the beginning of the study to collect only 1 percent or 2 percent of their patients. So they chose only the top 1 or 2 percent. After they did that, they looked at the population they had, and if any of the patients were already in a case management or disease management program, they were eliminated from the study, and then they simply instituted a new case management program with approximately 100 new members with two case managers.

Some of the issues that came up: They're one year into this, and they've not yet figured out if what they've done is of value. They know they identified members, they know that they're intervening with them, but everybody's wondering if it was worth the effort or the money. One of the products they used to do this was a neural net products that shall remain nameless. But one of the things that came up over and over again when they were working with case managers, RNs and the physicians was that the neural net was a black box to these people, and having worked with lots of physicians, something that's in a black box or done elsewhere won't be accepted by them. So for identifying these kinds of patients for case management, that's probably not the kind of product you want to use to do it.

The other thing that came up, and I know this is familiar to some of you in the room, is how do you measure a nonevent from intervention? So, for example, if the case manager intervened with a high-risk patient who had had four admissions last year for congestive heart failure and then this year they had one admission or two admissions or no admissions, did that case management intervention really cause that impact or didn't it? How do you count it, how do you measure it, how do you quantify the savings from that issue? That was a big deal to them, because they hadn't thought about that issue ahead of time.

The other thing that's obvious is that with any of these products you're going to have a false positive. So they took these 100 patients and identified them, put them in this case management program, but even with the best of data, 30 of them really didn't need to be in the program. So any time you use this risk scoring to identify high-risk patients, you're going to have some false positives.

Now, a couple of the tools you probably are aware of: you can set sensitivity and specificity on, so you can set the false positive rate any way you want it, but that's a major issue. The other issue is documenting return for upper management. This particular health plan hadn't thought through a methodology or a way to define their savings from this case management program. They also found out, and some of you have also experienced this and I know are aware of it, but all these tools seem to underpredict very high utilizers. So if you apply a risk score to a member, they seem not to be high enough numerically for the very high utilizers, and if you read about how these tools have been used, that's a repeated finding by most people who have used these. It's obvious to most of you that they will miss the catastrophic cases, but if you haven't thought through that ahead of time, and

Theresa referred to that from a different perspective, you'll miss what you're trying to accomplish mathematically. This particular group is moving forward into the second year. They're trying to define a savings methodology especially for the nonevents after intervention by case management and trying to identify some outcomes for this population, what they're really trying to accomplish with this new population they identified. They're also struggling with: What are the things you can really measure for these patients? And because of the problem with the neural net issue, they're thinking about using a different tool going forward.

So, thinking ahead, how do I see some of these tools being used organizationally? I think the pharmacy risk groupers that are out there now, and there are at least three of the vendors that are on that last list that have pharmacy risk groupers, I think they're going to be used a lot. They can't be used as much for some issues in medical management because all you're basing them on is pharmacy claims, but if you've followed this in the literature, they seemed to have equal predictive power with the diagnostic groupers. They're claims only. They're quicker, they're infinitely valuable because there aren't as many encounter claims available in many places in California. For most vendors this is a less expensive product. Some of these products are very expensive, and for some of you who don't have access to lots of capital, you might want to look at these because they're less expensive.

I think the organizations that have figured out how to do this are using a multi-department approach to look at these issues. I'm mostly talking about integration of clinical and financial or actuarial organizations. I think if you get one of these tools, you need to understand what you're trying to accomplish, what you want to measure before you buy the tool. Some people buy the tool and then decide what they're going to do with it.

This return on investment (ROI) methodology, I think, is important especially on the medical management side, because sooner or later the question comes up in most organizations, is there really any value in what these people are doing in the case and disease management part of the organization? Vendor selection: I think the organizations that are doing this right are using multiteams. They're getting references. I mean, if you're going to buy a product, call the people who have used the product; and we like to see them use the same tool throughout the organization. One of the things that these new tools are available for is to identify higher-risk members where improvement in health status is possible. Some of these have been able to identify patients in populations who have a diagnosis who aren't on medication, and that's really a new tool to identify those silent patients in a population who are at risk.

Some of you may have read about this, but one of the things that you can do with these tools now is identify patients at different stages of disease, and most of the tools have, for example, different episode groupings or risk scores for different types of diabetics. Well, as you trend the cost of what happens in these different subsets of diabetics over the years, we've already shown in a couple of places that

some of the diabetic patients who are identified earlier in their disease—I mean, this is intuitively obvious—if you identify these patients early and intervene with them, you can have a much greater impact both on the cost of their illness down the road and on their disease process. Many of the disease management programs in the past obviously were focused on the high-dollar patient, and so I think you're going to be able to identify patient groups earlier in disease processes where you make an impact.

One employer group identified some high-risk patients. These are patients with high-cost certain diagnosis and had no pharmacy refill, but because some of the episode tools look at all the diagnoses that the patient has, all the pharmacy they have, all the lab they have, etc., they can identify those patients who aren't getting pharmacy refills and intervene with them, and this is an employer group who identified these patients.

I talked about this earlier, but I think these episode tools that group patients by diagnosis to physicians who see this as correct diagnosis groupings, when you use that kind of tool to look at practice patterns, it's going to impact variations, because physicians just identify with this way to group patients. I talked earlier about this, but I think the pharmacy groupers are going to be used significantly more just because they're quick, easy and inexpensive.

MR. RON BECKER: Can you offer any ideas that would be a proper ROI technique? We're very concerned about using a vendor who does case management, and how much of their effect is regression to mean versus real effects from them? Do you have any ideas on that?

DR. LILIEDAHL: I'm not sure I can do that off the top of my head, but when I've been involved in that, we've done it as a joint clinical actuarial or financial exercise and actually had people in the room who were stakeholders in this issue and developed a methodology proactively with buy-in from the people who had to accept the results. I don't think there's one way to do it. I think you have to look at what data are available, the people you're working with, and come up with a methodology together. It's not a very specific answer, but that's a really big problem, a really big question that many people are struggling with.

MR. AXENE: I'm going to talk about some of the underwriting applications and what's done outside the clinical area, what's done outside of what's really going on today, and although you've had a reasonable introduction to it from our two previous speakers, I'm going to try to talk just a little bit about that and go through some of the applications and a case study on underwriting. I'd like to talk about is the future, because nobody can criticize me on that because it hasn't happened.

Just a little bit about predictive modeling. If you think of predictive modeling as a robust claims data-driven process, it's data mining to its extreme, to try and figure out who's going to get sick in the future and who's going to get really sick in the future, we hopefully have an objective process that is not subject to our whims and

our subjectivities in what we do. But basically the ones that I find to work the best are the ones that fit observed claim patterns to episodes of care patterns. In previous sessions at this meeting, there have been some really good discussions about episodes of care and how important that is, and I cannot emphasize how important that is in truly understanding predictive modeling. It's an emerging science. There's a lot of stuff happening out there. Those of us who have been in the actuarial profession for quite a while are extremely frustrated that the actuaries haven't led this effort, because essentially this is an area where we as actuaries are going to have to grab on to it in order to stay involved in this process. Right now it's predominantly a Ph.D. business, not an FSA business, and those of you who care about stuff like that, you might want to figure out what's going on.

The medical management applications are the most common ones today. Dick talked about that. You have the multiple disease management companies, you have the multiple risk companies, risk assessment companies and the predictive modeling companies, but basically the vast portion of the market today focuses on medical management. Underwriting applications are emerging. The actuaries have essentially had a control of the underwriting process, and now we have our friends in the more academic circles edging in to the underwriting process, and so, not that we're going to have to be defensive on this as actuaries, but you need to know that the applications that are growing the fastest today seem to be in the underwriting sector. I've done quite a lot of work in this area, and those who are doing it are still called early adopters. So if you're not doing it, you're not a late adopter. It's still the people who are doing it I view as early adopters, but very few are really doing this thoroughly as far as I can tell.

It's very controversial. First of all, you're looking at people with specific health conditions. Blacklisting is something it's been accused of doing. It's not really trying to do that, but the idea of trying to go out there and get rid of people with certain diseases—that's not what predictive modeling is all about. Some of our more aggressive peers might be trying to use it for that. There's a funny thing about HIPAA that you have to find out who has it in order to identify them to start working with them, and so there are HIPAA confidentiality issues. Sometimes, as the *Wall Street Journal* called it about a year ago with quotes from several actuaries, it's underwriting at time of claim. It's really not intended to be that, but taken the wrong way you could actually pursue that.

Many of you and many of your peers who are not in this room believe it's before its time, and others have told me it's nothing but smoke and mirrors. And so what you have is from an actuarial perspective, it's something that supposedly isn't here yet, yet on the clinical side there's been significant use, significant value, and I for one believe that there's tremendous value to the actuarial side of it, it's just that most of us are a little bit behind in getting there.

The most mature use of predictive modeling is on the medical management side where you're prioritizing your resources and focusing on those where you think you

can get the biggest bang for the buck. You help identify people who need particular disease care or disease management process, and frankly you're trying to get a competitive edge. If you can control that, perhaps your cost will go down, perhaps your premiums will go down, perhaps you can dominate the marketplace. One part of it that perhaps is ignored is that you're actually increasing the quality of care while you're doing that, because you're taking care of situations when they should happen, and when you do that, you're doing society a good favor; however, a lot of that is underplayed in the market today.

It can become a basis for risk contracting with disease management companies, because if you establish a benchmark, perhaps the disease management company will take on risk to produce that certain cost. Unfortunately there have been disease management companies that have gone out of business doing that, and there's frankly some health plans and some other companies who have gone out of business because they didn't have the guarantee delivered, and as a result they have to pay the bills anyway, so they also had problems. So it's not a panacea quite yet, but those doing a good job at it seemingly are very happy with the results.

Let's talk about underwriting for a second, because that's basically where most of my comments are going to be addressed. It's a very recent application of this. Who here uses predictive modeling on the underwriting side right now? How many of you did that more than two years ago? I don't see any hands. The small group areas are the ones that used it the longest. I remember hearing Howard Bolnick talk about this. The applications that I'm seeing today are in perhaps 50 to maybe 500 life groups. I haven't seen that longer than about 18 months ago. Now, there may be somebody that has an application that has been doing it longer than that, but those truly were the early adopters that did that.

I find it to be predominantly a play in the small- and medium-sized groups. Now, how many know what Group Health is really about? I used to be a member there for 25 years, so I know what it's about. They really care about people, and they have this philosophic bent that you don't do bad things to good people, and I can see why they don't want to do it in a small group because it's very volatile, but there are some players who feel less guilty about that, and they're out there doing it in the small group marketplace. But essentially it seems to be a small and medium market play right now. Large group: things tend to average out in the large group. Maybe if you divvy up the large group into segments or whatever, you'll find some emerging things, but I don't see a lot of people using this in the underwriting on large groups as of today. I do predict that down the road we're going to see a lot more for the large group because there's other value that you bring when you do this analysis.

It is more often a renewal-only application, because you're analyzing detailed claims data, and frankly, I don't know too many of you who will give your detailed claims data to your competitors. So as a result it tends to be focused on renewals, and

it's focused on understanding your renewals better than you've ever done before, so that you can decide what to do with your renewals. Now, let's imagine a situation where you have company A that knows a lot about this group, and company B that would really like to have this group. Well, if you're doing predictive modeling, you can do some strategic things. If it's better than average, you can perhaps be more aggressive on the quote. If it's worse than average, you can perhaps be more aggressive on the quote the other way. What's happening is once you understand your groups better, you can position yourself to understand your portfolio better, preserve it better, maintain the risk better, and so it adds all kinds of value to you.

Now, those of you who are traditionalists are going to view this very suspiciously because it's different. It's a change from the way you are used to doing things. It's far more detailed, far more onerous, and frankly you have not been very warm receivers of these ideas as I or others have come to talk to you about it, because, no, we don't do it that way, we have no interest in it. Vendors have been more focused on medical management and are now starting to focus more on underwriting applications. The long list that Dick had, and there were about 20 companies on your list, if you go through those, about half of them have started some type of activity that is very much underwriting/actuarially oriented. And what we're finding is more and more of them are developing that because they see this as a new opportunity, but the idea of applying it in an underwriting setting, I think, has some extreme benefits if you look at this.

Now, my theory on this is that it's an early adopter benefit. It's one of these things that if you are an early adopter, you'll have a one-time significant benefit in doing it. If you're a late adopter, you're going to be trying to play catch up, and I'm not sure that the benefit is there long term if you don't get some of the early adopter benefits of that, because, frankly, it's a game between you and your competitors. And if you can get the advantage before they do it, or if you can at least offensively take action against them rather than defensive action, I think that you will probably end up with more value. Now, the bad news is if the value is that slim, to be out on a leading, bleeding edge to do this, is it really worth doing it? And that's frankly where the companies are jockeying for a position. They really don't know. Most of the major players in the marketplace are using it to some extreme, I mean, to various levels of intensity, but frankly there are many that are just sitting there watching what's going by.

I have an underwriting case study here I'd like to walk through with you to show you how somebody could have or did or whatever look at this:

Task 1: Review current underwriting process.
Task 2: Present proposed predictive modeling process.
Task 3: Present information on predictive modeling options and preliminary recommendation for specific model.
Task 4: Select and license model.

Task 5: Install predictive model and complete initial testing.
Task 6: Implement predictive modeling into underwriting process.
Task 7: Parallel testing.
Task 8: Go live.
Task 9: Check back.

This is actually from a real, live example. In order to do this, the first step is to really understand your underwriting process. If you don't know what you do today in your underwriting process, there's no point in trying to introduce this. You need to meet with the staff to understand that, and frankly there's a lot of stuff that goes on in underwriting that hasn't been written down, so it's really important to understand that. The second step is to go through and present the predictive modeling process to make sure that you know what it is and help them pick the right option. I have never worked for an organization that has produced a predictive model, so I personally have no bias with what one should be because I own or I want to own it or whatever, but I'm finding that a lot of people really don't understand the advantages and disadvantages of the various models out there, and frankly oftentimes choose the last person that they talk to.

Selective licensing to model: as Dick said, the prices are widely different between the various models. They're starting to get closer together, because I think the dust is settling and people are going to figure that out. Some are annual leases, some are per-use leases, some are member based. There's a variety of licensing, but as you go through the process and install the predictive model, implement predictive modeling into the underwriting process. Task seven, which is the parallel testing, I believe is very, very important and sometimes overlooked. You need to integrate predictive modeling seamlessly in the process, or else it's just another step that you go along. Now, take an underwriter who looks at three years of experience, and he or she sees two high years and one low year, what is he or she going to do? Are they going to pick the high or the low? How are they going to do it? So sometimes there's credibility, sometimes you band them all together, sometimes you compare it to your rating basis, but when you have two high and one low or two low and one high, you're oftentimes trying to figure out Should I go high, should I go low, what judgment am I going to apply in this situation?

Take that two good and one bad or vice versa and now throw in predictive modeling. You now have another benchmark, another point that you have to consider in that process. The thing that kills predictive modeling in underwriting is subjectivity. You need an objective, consistent application of the process to make sure that it works. To replicate the scientific method means you have to be able to do it in an independent location and get the same answer, or else it doesn't satisfy the scientific method. Well, some of the applications I've seen on underwriting are about as arbitrary as having no underwriting rules, and it's just another feature that the underwriter looks at, sort of like the temperature or blood pressure of a human body, and I believe that it should be implemented in the process seamlessly so that it just becomes another part of the process that is systematically incorporated.

Well, parallel testing is important because you have to go back and see if it did any better. Are your rates higher, are your margins lower, are you making more money, are you losing more money? I think that parallel testing really is important, and I recommend at least a three- or six-month period of time to make sure you really want to do this.

I'm aware of some people who did that and pulled the plug; it didn't add anything. I'm aware of others who thought it was a tremendous boom, and so they went ahead and did it. But without that parallel testing, I think that you're walking into a garage without any lights on, and you might trip over something. Then go live and then come back and check it out to make sure that it's working. It's basically a systematic continuous quality improvement, you know, plan, do, check, act, and as you go through the process, I find that that works reasonably well.

In terms of the future, what are some of the other applications that are emerging or could be used? One of them is evaluating provider reimbursement levels as they relate to capitation or health budgets, etc. It turns out that when you don't do appropriate risk adjustment, you can miss the boat in doing actual-to-expected testing. Your budget may be 1.0, but you have costs that are running 1.1. Maybe the risk adjustor is 1.1, and you were right on. If you don't do risk adjustments, you blow your comparisons and you give faulty feedback information. The same is true of predictive modeling. If you have variations because of a greater or lesser percentage of the high-risk population, this is another way, sort of like reinsurance pooling or whatever, that can be used in any kind of comparison and development of capitation rates, health budgets, forecasts, etc.

I personally believe in multiple option pricing, where you end up with either a PPO against an HMO or a variety of different programs competing with each other. Understanding selection bias: this is the Holy Grail. I think that it really helps identify selection bias about as well as anything I've ever seen, and frankly this is what we try to use in understanding selection bias, and we find it to be a very effective tool. For incurred but not reported claims (IBNR), for the recent period, as I like to call it, or the most recent three months where you have no idea what is going on, it turns out that predictive modeling gives you an edge in understanding whether that blip in claims is real or not real, and so by using predictive modeling I believe that there's an enhancement to the IBNR process in the recent months to get a better handle on where your experience is going. In talking to some of my friends in the industry, they view the IBNR process as the most strategic part of their organization. Now, I never really enjoyed doing them, but I understand that it gives you the earliest forecast of where you're going and trying to figure out where you're headed, and if you can add a bell and whistle to that to help you do a better job of those most recent months, you may get a one week, one month, six week head start on understanding the market better than perhaps your competitors. This is an application that I think is natural for the actuaries, and I'm encouraging people to take a look at that.

An area I'm just fascinated about is consumer behavior modeling, and this is something that has not hit the health care sector as much as I would like to see. Our property and casualty (P&C) friends are way, way ahead of us in understanding this. I don't know if there are any life people in here or not, and maybe the life people are into this too, but the P&C people have made significant headway into understanding customer behavior or consumer behavior. Now, there have been several sessions here on consumer-driven health plans. I'm not talking about consumer-driven health plans, although that's part of the process. I'm talking about consumer behavior and how you consume health care, and it turns out that using predictive modeling helps in understanding where things have gone and in understanding characteristics of the people who were having those claims. In trying to understand and, for example, linking up and correlating the complaint logs at the doctor's office or through your customer service department—linking that up with your claims pattern gets you some very interesting results. It turns out that the highest costs come from the people who complain the most, usually in most plans. In addition, the highest claims are the ones that you can manage the most, and perhaps you can improve your complaint ratios, which would lower costs by using predictive modeling to better understand why they're having high claims. And so if you can get people in advance who are having high claims, you may be able to actually resist several factors that are impacting your health care trend just by understanding consumer behavior.

The mere linking of your complaint logs from customer service gives you an edge there. There are other aspects about buying habits through the Internet and whatever else. It's harder with HIPAA today, but basically there's tons of information you can buy that is Web based, that if you choose to link and correlate it with your claims data, it's amazing the amount of information that you can use to understand the behavior patterns of people consuming your health care services. Predictive modeling is right at the core of all of that, and it's one of the best tools that I know because it gives you the opportunity to better understand your people.

Predictive modeling is also beneficial for pricing theory applications. You're introducing a new product, let's say, and you're trying to understand where the costs for that new product are going. Perhaps it's a consumer-driven health care product or whatever, but if you better understand the characteristics of the people consuming claims by looking at their episodes, and then once you have an episode-based model, understanding your distribution of those episodes historically, you can actually come up with better methodology of pricing, whether it's elasticity, and I'm trying to figure out how high your price can be and they'll still buy it, or what you can do to channel them to a program that maybe you would like them to go through, or whatever. I believe it enhances the applications there and just opens up new ideas for you to pursue more sophisticated pricing approaches. We've seen this work reasonably well in trying to gauge the prices for consumer-driven health care products. That's one natural thing.

Anybody in this room know Ken Aruda? He used to be at Blue Cross, Blue Shield of

Maryland. Ken was a client of mine many years ago, and he introduced me to a word called psephology—and this is where I give Ken credit for the word—but psephology is based on the Greek word *psephos*, and in old Grecian times they used to vote by putting rocks in big urns. The candidate that had the fullest bucket won the election. Back then they had a science called psephology. Today we call it exit polling. Basically they had a science of the study of voting patterns. Well, the science of how you choose health care benefits today between this plan or that plan is psephology. It fits. It sounds fancy, and you can sell it as a consultant. Anyway, what happens is how people vote with their feet on different health plans in a consumer-driven health care program, pricing theory, that's what it's all about, and I believe that predictive modeling is one of the tools that helps you get into that, so that you can do it right, because the key is predicting where they're going and seeing if you want to take extra risk.

MR. STEVEN DUNCAN: I have a question for Dr. Liliedahl. I used to run a company, and what we ran into with health plans in the medical management area was their belief that they already knew all their most risky cases, because the highest predictor of future behavior is part behavior and part hospitalization. So one of the things we did was to sort of tease out, to see whether we could build models that use variables that did not involve hospitalization. So we were trying to predict what people had not had a past hospitalization, and the best we could do at the highest level was about a 40 percent probability of an admission, whereas if you take the people who have had the higher past hospitalization, you're up in the 80 and 90 percent probabilities. So it became very difficult to sell an expensive implementation of something that really didn't add or added only very marginal value to their existing knowledge. My other comment has to do with ROI. I'd be interested in the feelings of the group about the notion of doing controlled tests. We always insisted with our clients that they in fact run randomized controlled tests, and that was my experience with our clients. But since then I've found a lot of opposition, a lot of resistance to the notion of randomization, and yet at the end of the day it's still probably the best way that we have of demonstrating value.

FROM THE FLOOR: In the SOA's study on predictive modeling that Milliman put together a few years ago, one of the models was the University of California in San Diego's model, which I looked into. It's in the public domain, and I'm just wondering if you guys have any experience with that and how it compares to commercial models. Any comments on that model?

MR. AXENE: That's primarily a Medicaid model. You can all access it now at no cost, to my understanding. It also has a pharmacy model attached to it. It has not to my knowledge been created for a commercial or Medicare population.

FROM THE FLOOR: What I think they said in the Society study, though, is you can recalibrate it for the commercial market, and the predictive value wasn't quite as high.

MR. AXENE: It is mostly a risk-adjustor type model, and it's not necessarily predictive modeling in terms of predicting just the overall sick people. It is a highly used model, and Dr. Kronick is the one at UCSD who actually built the model. It's been used widely in Medicaid state insurance departments and is actually one of the more competitive models. It is being adopted by some of the commercial players, but it gets labeled as a Medicaid model, and even his latest adaptation of that, a more comprehensive version, is still being "labeled." The beauty of it is it's an open box. He developed it as a public service basically. Other ones are probably more popular and readily able to use. I believe one of the shortcomings of his model is the clinical relevance to changing physician behavior—and both Dick and I think it's very important to have something that's clinically relevant—which gets you to understand episodes of care or treatment patterns. That's something that's really important, but it's a very good model, it was included in the study. There were several that weren't included in that study that are also very good. Either they weren't asked, or they didn't get the data in time.

MR. DOUGLAS MCCANN: I work for Highmark, and I manage one of their underwriting areas for mid- and large markets. I was looking at the underwriting case study that you had up there. Was that a proposal or was that an actual case study?

MR. AXENE: That was extracted from a proposal; however, it was actually used with one particular client.

MR. MCCANN: I guess I was concerned that we got to step seven before there was parallel testing, and in terms of managing an underwriting area, in following a new or different or modified way of doing the pricing, it would be very disruptive into what is usually a very high-pressure situation. I was wondering, was there a testing and a development of the model or a proving out of it before that, or did we actually get to step seven before we parallel tested it?

MR. AXENE: Well, the parallel testing was after it was up and running and seamlessly integrated in the system. There were intermediate testing steps upfront, but as it was finally ready to go, it was basically dual work for a while where people did both ways what they used to do, versus what they did do for a period of time, and that parallel testing was basically the last check before it went live. There were other tests and calibrations that occurred all the way through the process, but at that point it was, "Let's make sure, and that test turned out in every application I'm aware of." That was very important because of the things that came up that nobody had anticipated.

MR. THOMAS DOREN: I wanted to talk about one of the comments that was made about the tendencies with these models to underreport catastrophic or large claims. Is that indicative of all the models, or is any one of them better than others at catastrophic?

MS. KEANE-NORTH: Many of the models are calibrated using a truncation methodology, so they're actually designed to not predict the high-cost cases, but I would suspect that there are those out there that don't use truncation. I'm not familiar with any of them quite honestly.

DR. LILIEDAHL: My experience is it's sort of like the old cost assessment spiral, that as you try to raise the price to take care of the few remaining people, you often underestimate that, and I think that there's always somebody who's going to be beyond it. So I think that in the extreme it will always underestimate something, plus unless you have infinite categories, there's always somebody who's been grouped with somebody else that's high and low combined. But I would say some of the methods are worse than others at trying to predict the catastrophic, although all tend to understate on an individual basis.

MR. AXENE: When I made that remark, I was primarily talking about medical management and identifying patients for case management. What I was trying to say when I was talking about underpredicting severity or high-dollar patients who aren't catastrophic, if a model predicts a risk or I just ask Theresa what the highest risk score was for the tool they use for a patient predicting forward, and it's around 30, which is 30 times a multiple of one, what I was trying to say was when you look at populations from a medical management perspective, if they identify a risk of 30, they have patients who have costs they incur in the next year that are not 30 times one, but 35 or 40 times one, and that seems to be true of most of the tools that I'm familiar with, at least on the medical management side. I don't know if that's any more clear, and I don't know of one that's better than the others on that issue.

FROM THE FLOOR: I'm curious what the market receptivity has been in underwriting applications. I think everyone understands the age and sex rating. Everyone in the broker community or consultant community understands experience rating, but what are the conversations like with large groups around "your average group risk score is X and here's how we calculated it?" What kind of information do you have to provide? How do you justify that?

MS. KEANE-NORTH: We actually have found that most of the large employer groups that we've had discussions with are very receptive to this, but we have to provide a lot of really detailed information. We have to, number one, explain the model so that they understand how it's working, and then, number two, tell them which disease categories that they seem to be high or low in so that they could gain a really good understanding of what this is, and as a result of that, I think that we've seen a lot more demand. We actually have a purchaser advisory council at Group Health that has 12 members; we gave these 12 employer groups their risk reports, and they are just eating it up and are asking for more, so there is lots of really good receptivity.

DR. LILIEDAHL: I don't work in the large employer market, I only basically work

with health plans and health systems, so I hear stories secondhand. But I would say that there is reasonable receptivity, especially if you can compare it to a norm or a benchmark. People just don't like a number. I mean, they hear 1.1, they don't like that number, but if you say you're 10 percent worse or 10 percent higher than a group of your size or a group of your industry or whatever, comparing it to something increases the acceptability of it. Now, from a totally different perspective, the receptivity of health plans to get into this and do this underwriting—I would say it's an early adopter world, and it's not that friendly yet.

MS. SABRINA PHELPS: We at Blue Cross of Louisiana do use predictive modeling tools, and I find there's a big difference between the way the large insured groups want to talk about this and the way the large self-insured groups want to talk about this. The insured groups tend to look at it as another arrow in the insurance company's quiver to increase rates without them understanding it, but the ASO groups look at it as a tool to help them understand what their costs are going to be, and that you really are working to try to figure that out for them and do something about it. So that's a big dynamic there, and we've tried to get in front of our ASO accounts a little bit more than our insured accounts.

MR. AXENE: Do you want to share any of your predictive modeling experiences?

MS. PHELPS: Well, we are just getting into it. We started in the medical management area as a case management referral source, because we didn't really have a good referral mechanism. And actually, our first test was we followed our case management list of who was in case management the last year, and then we ran our predictive modeling tool on the data of a year ago. We said, okay, what's the correlation of who we were getting versus what this tool would have given us, and it was pretty good. That kind of goes with somebody's comment about our referral mechanisms are good, but there were some people we weren't catching, and those were our chronic members who really weren't incurring a lot of costs but were on a lot of drugs, and we finally had the drug picture in. One of the reasons I really pushed the investment into this was because our case managers really didn't have what I'll call a virtual medical record of the member. We had medical claims over here, which was a bunch of codes on the screen that only a customer service rep might understand, and the tool we purchased was not just a mathematical tool, but it presented the data to a case manager so they could see what was driving the score, what contributed to the buildup of the score, and then there was another little view that showed what drugs they were on. So the case managers had information to get on the phone with a member or a physician to say, "We know what's going on with this member. We didn't have that before. We are exploring right now putting it in our small group rating formula as just another variable in our tiering formula; we're not going 100 percent. We're probably going to weight that very low in the beginning and move that weight up as we see that it's helping our retention of our tier one groups."