RECORD, Volume 31, No. 2*

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

Session 80 Seminar
Financing Chronic Care Seminar: The Costs of Chronic vs. Acute Conditions

Track: Health

Moderator: CHARLES S. FUHRER

Panelists: MICHAEL COUSINS†
IAN G. DUNCAN
NANCY A. GARRETT‡
IVER JUSTER††
JAAN SIDOROV‡‡

Summary: This session takes a fresh look at the issues related to "acute care focus" and the challenges of providing coordinated care for chronic conditions in the United States. This three-part seminar encourages a deeper exploration of this topic from a multidisciplinary perspective. Some of the questions examined include: How do you predict the incentives of costs for an individual with a chronic disease? What are current obstacles that stand in the way of health-care system improvements for chronic illnesses? What is the impact of the aging U.S. population on the chronic and/or expensive conditions most often included in disease management programs

†Dr. Michael Cousins, not a member of the sponsoring organizations, is senior vice president in research and development at Health Dialog Analytical Solutions in Portland, ME.

‡Dr. Nancy A. Garrett, not a member of the sponsoring organizations, is in health service and reporting at HealthPartners in Minneapolis, MN.

††Dr. Iver Juster, not a member of the sponsoring organizations, is vice president in outcomes management at Active Health Management Inc. in Sausalito, CA.

‡‡Dr, Jaan Sidorov, not a member of the sponsoring organizations, is medical director in care coordination at Geisinger Health Plan in Danville, PA.

(e.g., heart disease, diabetes, asthma, psychiatry and chemical dependency)? What is the comparison of health-care cost trends between chronic and non-chronic condition members of a health plans? Is there a prevalence of chronic conditions within the insured population, and what is the influence of possible confounding factors (such as catastrophic claims) on trends? Attendees gain a greater understanding of the challenges presented by the prevalence and trend of chronic conditions in the United States, as well as the benefits of improving the ways in which these conditions are managed within the health-care system.

MR. CHARLES S. FUHRER: We're going to be dealing with the cost of chronic care. We have a number of very good speakers. I'm with the Segal Company. Our first speaker is Michael Cousins. He works for Health Dialog Analytic Solutions in Portland, Maine. He has a Ph.D. in neuroscience and has been working on predictive modeling and health outcome evaluations. This is the second time that he has spoken at a Society of Actuaries meeting.

DR. MICHAEL COUSINS: As Chuck said, I started as a research scientist in neuroscience. I studied Parkinson's disease and the real biological basis for neural networks back then. What struck me as most interesting was the methodology, and that interest has continued over the last decade or so. Prior to joining Health Dialog in its data research center, Health Dialog Analytic Solutions, I headed up the research department at another disease management (DM) company. Now I've been able to get more and more into research in data analysis at Health Dialog's Analytic Solutions. It's a subsidiary that just focuses on data analysis, so for people like me or us in this room, it's an ideal world.

Before I start, I want to mention a colleague of mine, Scott Pollard, who is the one that's responsible for all the good stuff in this presentation. I'd like to frame today's talk. This is not a "feel good" presentation. I'll probably be raising more questions than answering them. My goal is to provide you information that I learned about and that I'm interpreting in a particular way. I'd like to share that with you, and perhaps you'll see it the same way, perhaps not. But either way, I'd like to make you aware of this information. Some of this information you've seen over the past couple of years. As Chuck said, I spoke with my esteemed colleague John Stark before at the Society of Actuaries, and Ian has spoken somewhat about what I'll be talking about today. However, I'm going to be framing it a little differently today.

What we're trying to do in DM is lower claims costs, improve productivity and so on. The way we're trying to get there is by improving health quality and status outcomes and by optimizing utilization. We don't get to this end by denying benefits or changing benefits. In this evaluation side, what we're trying to do is see the relationship between what we think we're doing, with our nurses, pieces of mail or information that goes out and outcomes. What we're trying to do is look at cause-and-effect relationships. We are looking for causality. This is a point that a colleague of mine, Tom Wilson in the Disease Management Association of America (DMAA), has been driving home over the past several years.

All of the measurement methodologies that are out there, including the ones I'll be talking about today and that we've heard about before, have pros and cons and strengths and weaknesses. It's important to keep these in mind so that when you're looking at results, think about what method was used to get to these results and realize that no matter what the methodology is, there are important factors that need to be kept in mind when interpreting the particular results.

My view on the world of methodologies is essentially that there are good ones for establishing causality and there are good ones that address practical concerns, such as time and financial constraints. I think of these two traits, causality and practicality, on a continuum where we can fit each of these methodologies. In the one corner, weighing in as the gold standard, we have the randomized control trial. That's really good at establishing causality but, in a lot of situations, it's not terribly practical. In the other corner, we have the pre-versus-post, also known as a historical control design. It's very practical and easy to implement. It's a before-and-after comparison, but in terms of establishing causality, it's not so good. In the middle, we have what are called quasi-experimental methodologies, including regression discontinuity, propensity scoring and variations all over the place. Again, no matter what methodology we use, it's important to remember that there are going to be trade-offs between practicality and causality. No matter which one we use, there are going to be important methodological issues.

Now I'd like to focus a little on the pre-versus-post methodology. Why? Because it's the most widely used methodology in our environment today. It's used by commercial care management and DM programs. It's promoted by nationally known consultants and national consulting companies. Recently, Ian Duncan of the SOA has done some wonderful work and published a series of papers that have brought some clarity to how the pre-versus-post methodology should be applied when evaluating DM programs.

There are many pressures on us as researchers, evaluators and actuaries to develop an easy-to-use methodology in a commercial setting. The list of options isn't terribly long. Because of this, I think, the pre-versus-post methodology has been widely used. It's easy to use and doesn't require a lot in terms of expertise or conceptual understanding, but it does require a lot to pull all the data together, test it for completeness, etc., and then, most importantly, agree on the assumptions for the parameters that are going to be used to actually bring this methodology together.

I think it's important to define what we're talking about. Over the years there have been several different versions of pre-versus-post design. There's a cohort, where we're just tracking a group of people over time. There's methodology that was popularized at the turn of the century, the population-based pre-versus-post adjusted historical control, where they started to look at entire populations, not just track cohorts, and then apply cost and utilization trends. Most recently, we add on

to the advances over this turn-of-the-century time by having what we call incident and prevalent adjustments to help control for regression of the mean and to help adjust for a turn in the population that can magically produce savings even if the program has no effect.

As I said a moment ago, Ian and others such as myself have spent a lot of time and effort refining these pre-versus-post methodologies, trying to articulate clearly why it's important to keep in mind the limitations of the methodology and also trying to articulate exactly how to go about doing it, because it's not a straightforward endeavor. In a few minutes I'm going to be backtracking in a sense and saying that I don't think this is the way of the future.
I'd like to say my key messages now before actually getting into the data. The first one is to be aware that these adjustments, whether they're actuarial, cost and utilization trends or incident and prevalent adjustments, do not overcome the major methodological weaknesses of the pre-versus-post design. Methods matter. The method used, or the glasses we use to look at the numbers, impact the results more than any of us. Another key message is that no one number should stand by itself. Remember the DM value proposition, where we have financial and utilization and health quality and status outcomes? What I implore practitioners to do is not just look at cost. Don't just focus on that. If nothing else, at least look at the relationship between cost and utilization, utilization and health quality and status outcomes. Finally, view results with skepticism; use multiple metrics and methods, plausibility tests and so on. My take-home message today is going to be that because of the problems with the pre-versus-post method, the evolution that has gone on in my mind has been to focus on what we can measure accurately, which is service level or a particular activity that the DM program can deliver. I'll elaborate on this as we go through this.

Because of the 20-minute time frame, the data that I'll go through will be at a relatively high level. However, as I said earlier, John and I spoke about this before and our full-length presentation from last year is on the SOA's Web site from last year. We have a paper that describes this in more detail than I'll be covering today. Let me know if you're interested in this stuff.

What did we do? We looked at approximately 900,000 PPO members and then identified members with diabetes, coronary artery diseases (CADs), congestive heart failure (CHF), asthma and chronic obstructive pulmonary disease (COPD). Then we looked at some exclusion conditions. Essentially our strategy was to use the same DM program, the same population and the same data set and just segment the data different ways. The story is that the same data looked at different ways can produce tremendously different outcomes.

In this first study, we looked at two different ways of identifying a population. In the first we used what we called a "looser criteria," where members were identified with diabetes, CADs, etc., if they had one or more inpatient or emergency room visits and the diagnosis could be in any of the positions on the claim (the first,

second and third). In contrast, we had the second methodology where we only kept people who met the criteria if that diagnosis for diabetes, code 250, was in the first position. It had to be in the primary diagnosis. Relatively speaking, the second criterion is looser than the first. Again, sticking with the theme of take-home messages first, the criteria we used to identify the members have an impact on the results.

Here are the results that support the conclusion. We have two different criteria, the looser and the tighter. Frankly, the message isn't that we used the primary diagnosis or the first three positions. The message that I would like to get across is the fact that all it takes is changing that criteria. It doesn't matter what it was that we used exactly; it's the fact that these small differences can affect the results. With the looser criteria, we have $482 per diagnosed member per month (PDMPM) baseline trended and $426 PDMPM, so about an 11 percent savings. With the other criteria, we had about a 12 percent difference between baseline and year one. Although this seems like a small difference, when we're talking about gross savings of a DM program in the range of 1 percent to 5 percent, this is not all that insignificant. As you'll see in a second with the other criteria that we modified, the impact actually can be greater.

In the other little study that we did, we changed the way that the members are excluded. So in the program people are identified with diabetes, CAD and CHF and all the members are managed unless they're being sent off to a case management program, but then when it comes time to do the evaluation, we take out people who aren't expected to benefit from the program, either because they're institutionalized (they have HIV or cancer, etc.) or claims costs are volatile. Those are typically the two reasons we take out people with these exclusions. The criteria used to exclude members or claims from the analyses impact savings.

In this particular study, the impact was relatively small when we excluded members. We have that 11.5 percent savings again. Here it's 13.9 percent. This is only a 2 percent difference. This study was relatively small, but I do know, not just from this study but from the countless others that we've done where we've assessed the impact of exclusions, that we have differences based on whether or not we used exclusion criteria that ranged from 1 percent up to 9.9 percent. The point remains that it's not the specifics of the exclusion criteria that we're using; it's the fact that all it takes is to use them or not use them. As I said a moment ago, when we excluded claims instead of members, we got a different savings amount. It's that same story: change the assumptions or the parameters and we can change the results.

In this next study, instead of just looking at the chronic people, we now looked at the non-chronic, or the index population. We did that because that's the population that is typically looked at to construct the cost and utilization trend. There are two groups. In one we excluded the members (end-stage renal disease (ESRD), transplant, HIV, cancer, etc.), and in the other we kept everyone in (no exclusions).

The bottom line is that the criteria you use to exclude the members from the non-chronic index population impacts the trend. That's a circular statement in the sense that this is the data that I used to construct the trend, so depending on how those groups are constructed, we can either have big savings or small savings.

There's a 19 percent difference between those two. The per-member-per-year (PMPY) trend is 11.4 percent. With a baseline cost of $102, $113, from an index population, non-disease, no DM program, there's 11.4 percent savings. When we look at that index population and we don't take anyone out, there's a 9 percent PMPY trend. If we were to apply this 11 percent number to baseline and program year, we'd get one dollar amount; if we apply this 9 percent, we'd be getting yet another dollar amount. Again, although this may seem small (2.19, which is the difference divided by 11.44), that's a 19 percent difference. That's substantial. We're talking about a 19 percent difference depending upon whether or not we use exclusions in this index population. Of course, there are the same points I made earlier about whether or not we exclude members or claims or whether we take out transplants or not. The list goes on and on. We're going to be affecting the actual program results. What are the program results? The program results are the product of the glasses or the prism we're using to look at the results. That's a key message.

If you look at how the two groups track each other over time, say 12 months, it's parallel. They're not grossly different; one is shifted up. The exclusion-members-removed group is the higher one and corresponds to that 11.5 percent.
Because of the time limitation, this had to be a brief overview, but I wanted to point out that there are other sources, this as well as the work that Ian has done, for you to learn more if you're interested. Now I'm going to get to the home stretch here, which is back to causality.

There was an article in *USA Today* that came out from the Centers for Disease Control and Prevention (CDC) that said "Hospitalizations for Diabetes Fall." How does this headline differ from the headline that you see from your DM programs? It really doesn't. I've seen press releases, my own and from big companies like American Healthways, that are putting things out like this. The fact is, though, that we don't know what is driving these improvements. Who knows if this is because there's a critical mass of DM programs in place across the country? That's the big question. That's why I think it's so important for us to keep in mind that when we're using these pre-versus-post evaluation methodologies, we're not truly getting at answering the question about causality.

What are my conclusions? As far as the pre-versus-post and adjusted historical control, I never know which way to go with it, so I put them both sometimes. Including the incident and prevalent adjustments or any other actuarial adjustments are non-experimental. For health-services researchers such as myself, this means that causality cannot be clearly established. The CDC study is an example of that. In my judgment, the pre-versus-post methods are too sensitive as design

assumptions (for example, the different ID criteria, exclusions and so on). The bottom line, in terms of a conclusion, is that the high-resource requirements, meaning how much time and effort it takes to put these sorts of evaluations together, and the fact that it can lead to poor decision-making, outweigh the benefits of providing for the perception of rigor. I work for a subsidiary of a DM company, and we put out evaluations with the pre-versus-post method. Why? Because we are pressured to show the value of a DM program. My recommendation is to stop the madness and instead to generalize results from rigorous evaluations and randomized control studies. Believe it or not, there are some out there, the biggest of which is Medicare's Chronic Care Improvement Program (CCIP) coming up. If you don't want to wait for that, there are some others. Again, in our own company we put people in all these different blocks and do all the actuarial accounting so that we can have a fully transparent, rigorous pre-versus-post method, but the transparency doesn't address the critical issue of causality. The key message is that transparency does not address the critical issue of causality.

What is one to do? Don't rely on a pre-versus-post method to judge the success of care management programs. Instead, make judgments generalized from other more rigorous studies. Henry told everyone yesterday about PubMed. You can get these off of PubMed. Not all of them are favorable. DM is not a magic bullet. There needs to be a confluence of a lot of different things, such as the right population and the opportunities for improvement. If everyone is already at the top in terms of health, there's not going to be a whole lot of improvement. The interventions need to be cost-effective and effective. They need to be based on evidence-based research. Don't be falsely comforted by full disclosure; transparency does not address causality. Focus on what can be measured accurately. I recommend focusing on levels of service delivery, on the actual program activity. Put the contact guarantees on that. To do otherwise is frankly a waste of time.

But if, despite the empirical evidence and despite this very eloquent 20-minute talk, one must evaluate a DM program with a pre-versus-post method, I implore you to do two things. One is, look at more than one metric; don't just look at the DM outcomes. Go upstream. If nothing else, look for correlations. Also, use multiple methods. There are quasi-experimental methodologies, regression discontinuity and match control. There are also some in-between ones, such as looking at participants versus non-participants. Seeing a positive effect there is, in my view, necessary but not sufficient to show success of a DM program, but it's one other place to look. Or look at time course (not time series, but time course). Is the fact that someone is enrolled in the program longer correlated with a greater effect?
FROM THE FLOOR: Given all that you've just said, what do you think the CMS study is going to end up with?

DR. COUSINS: The CCIP 721? I was part of the team that designed that, so I'm not sure I can say anything bad about it. Actually it's a great design, and it's balancing the need for knowing the cause-and-effect relationships with the practical facts. What's a practical fact? Randomization at a personal level is probably not

terribly practical. So what have they done as the next best thing? Do the randomization at the provider office level. I think that when those results come in, whether they're good or bad, we're going to have a pretty good sense of whether or not these programs are effective. The short answer is that I think it's good.

MR. FUHRER: Our next presentation is a paper entitled "A Comparative Analysis of Chronic and Non-Chronic Insured Commercial Member Cost Trends." The authors are Rob Bachler, an FSA, of American Re HealthCare of Princeton, New Jersey; Ian Duncan, also an FSA, of Solucia Inc. in Hartford and Iver Juster, our presenter, who is at Active Health Management in New York City. Iver is a family physician who works in medical informatics and health outcome evaluations.

DR. IVER JUSTER: Having worked in the field of health economics, which is modeling what really happens, I've fooled people in my job into thinking that I'm actually measuring what really happens by looking at real claims. I now come to the conclusion that my next big product is going to be building a machine where you dial in the return on investment (ROI) that you need in order to win on your DM contract (we're a DM vendor), the assumptions that you think your customer is willing to take and the ranges that you think they're willing to believe, and you'll probably get five or six believable scenarios that will support that you did indeed break trend by 2 percent and therefore, they should pay you more money.

Let's continue with talking about the problem of measuring financial outcomes, not modeling them. The other half of what I do is to model them. I know that a heart attack costs so much money in a year, and if I give a bunch of people beta blockers that weren't on beta blockers but need them, they are somewhat less likely to have a heart attack. I just multiply numbers together and say that you ought to save this much money. This is about supposedly measuring these things rather than modeling them. We'll continue to talk about this cost trend as an estimator. In other words, we're going to continue to talk about, "The trend should have been 10 percent, but it was 9 percent. What does it mean by 'should have been' 10 percent, and should it really have been 10 percent? If it should have been 10 percent and it really was 9 percent, then our DM vending organization will applaud for us." We're going to look at "should it have been 10 percent" and what we mean by that. We'll look at some case studies that we did in order to help eliminate this issue.

The DMAA's definition of disease management is "...system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant. DM supports the physician or practitioner/patient relationship and plan of care. It emphasizes prevention of exacerbations and complications utilizing evidence-based practice guidelines and patient empowerment strategies." It's the first word that I want to emphasize here. It's a system. Generally speaking, health care is not a system, even though they call it the "health-care system." DM has tried to make some part of that health care, especially for people with chronic conditions, more systematic. In doing so, you bring evidence-based health care, in which everybody ought to believe, right? You

ought to give beta blockers to people with heart attacks unless they have a contraindication. So why aren't many of them getting it? From the evidence base that is from research, metanalysis, ultimately clinical trials or, at the very least, really good expert opinion, doctors and patients ought to do what's right to prevent clinical adverse events, like more heart attacks. Heart attacks cost money, so you ought to save money. That's the premise.

In any design, whether it's experimental or some sort of version of pre/post, we need to answer the question: What would have happened if we wouldn't have been there? If we wouldn't have given the beta blocker, how many heart attacks would there have been? If we wouldn't have DM, what would the cost trend have been? As Michael said, you can't just compare pre and post because if the people aren't the same, then you have a turnover problem. If they are the same, a bunch of sick people aren't nearly as likely to be sick next year, and you'll get regression to the mean and you'll look better even if you do nothing.

What you need is some sort of comparator group, reference group or control group—whatever you want to call it. There are two problems. In DM, even randomized controlled clinical trials have certain problems, mostly because DM is not a drug. It's not a statin. There are a lot of issues with using disease-control experiments on DM that there aren't in giving a drug. Nevertheless, it's better than anything else we have, such as pre/post, which is what you usually get because you're in a situation where all you have is the whole population. Sometimes you don't have some other population that doesn't get the intervention.

We're going to focus only on this pre/post; all you have is the whole population. Now what do you use as the reference group? As Michael said, what you usually use as the reference group are the people that didn't have the chronic disease. If your program only treats diabetes, then the reference group is everybody who didn't have diabetes. If it treats 10 diseases, it's everybody who doesn't have one of those 10 diseases. We will call the people who have one of the diseases that the program measures, the chronics, and we will call the people who don't have one of those diseases, the non-chronics, even though they might have a chronic disease anyway. The assumption that's often made, with various fancy adjustments, is that the trend of the chronics and the non-chronics would be the same if you didn't do anything to anybody, that is, if you didn't have a DM program. That's what I meant by "it would have been 10 percent." If the trend of the non-chronics is 10 percent, then we assume the trend of the chronics would have been 10 percent absent the intervention. The intervention is this system called DM. Is that really true?

One reason it might not be true that's important, but possibly trivial for this particular reason, is that you don't know what's going to happen next year, so you can't use last year's trend as this year's trend estimator. Maybe you can use the non-chronic trend from this year as the chronic trend estimator for this year. Let's do a simple example. Say the baseline cost was $6,000 PMPY for some group of chronic people. The non-chronic trend was 12 percent, so you predict that in the

absence of intervention it should have been $6,720. But it was only $6,300, so you saved $420 per person. If there were 20,000 person years, you made $8.4 million. If you paid less than $8.4 million, then you got a positive ROI. That's what's usually done.

There are a number of assumptions that we examined in our study (there are a few that we didn't examine). The major one was: When does somebody enter the chronic pool? In our study we used five diseases to be chronic. When was somebody chronic? Let's talk about heart failure. We didn't go out and measure everybody's ejection fraction every single day and the day that it fell below 45 percent classified them as having heart failure. Instead, what really happened was that the patient came into the doctor, short of breath, couldn't lie down to sleep, maybe with puffy ankles, and the doctor said, "There are a bunch of things that could be going on, but it could be heart failure." After all, he has to get paid, so he's going to put "heart failure" down, but in his head he's saying, "Rule out heart failure. I'm going to give the patient some diuretics and see if the patient does better." The next week, the patient comes in and says he or she is feeling better. Six months later, the doctor finally gets an ejection fraction, and nine months later, the patient gets started on an ACE inhibitor because this isn't a health-care system and the doctor didn't think of putting the patient on an angiotensin-converting enzyme (ACE) inhibitor. Eventually, if you look at the claims, you build up the picture that this person really had heart failure. Now, some of those people didn't really have heart failure. They had something else mimicking heart failure and eventually the doctor found out what that was and changed the codes, but now they're forever coded as having heart failure for some defined period of time.

What do you do with these people? When do they enter the chronic pool? Remember that we're using the non-chronic trend as an estimator. Well, it could have been entered prospectively. That is, whatever your criteria were, whenever they fulfilled that criteria, say January 3, 2005, that's when you called them "heart failure" and you classified them as having heart failure forever. That's the prospective method. In the retrospective method, they're tagged as being chronic all the way back to the beginning of whatever you're measuring, under the assumption that maybe they had heart failure forever or at least in the few years that you're measuring. We call that the "ever/never." If you ever had heart failure, you're always classified, and if you never had it, you're always in the non-chronics. The other side of the question is: How do they stay in the chronic pool? In our particular set of data, we said that once you're chronic, you're always chronic. But you could have to re-qualify every year. Maybe you had those two claims for heart failure in year one. In year two, the doctor found out that you didn't really have heart failure. Should you really be called somebody who has heart failure that year?

In our study, we compared chronic and non-chronic trends over time, in this case four years, in an untouched, commercially insured population. We varied the following assumptions: how the chronic population was selected (that is, how did

you enter the pool of chronics?), adjustment for service mix (that is, of the amount of money that was spent, for everybody's pie, how much of my pie was spent in the hospital, for drugs, out of the hospital, for x-rays and so on?) and an adjustment for risk (I have heart failure, but Bob over here has heart failure, diabetes, hypertension and peptic ulcer disease. His cost trend may be different from mine because I don't have all those co-morbidities). The study population was 1.5 million lives. We imposed no minimum eligibility criteria; you could be there for a month (although there wasn't a lot of turnover in this particular data set). If you had a claim for $73 for some month when we couldn't find you as a member, the $73 was tossed out; it didn't enter the numerator and you didn't enter the denominator for that month.

The first study we did was prospective (once chronic, always chronic). From the day I'm classified as chronic, I remain chronic for the rest of the four years, but I'm not classified as chronic before then. The years 1999 to 2002 were the four years. In this scenario, the chronic population accounted for 4.1 percent of the members, but 14.3 percent of the cost. That's not so shocking, but in the next four years, the chronics were now 8.6 percent, more than doubled, and they didn't quite double their costs, which were 23.1 percent. Is this prevalence creep? Just in the general population, I think the prevalence of chronic diseases is going up, but did it double in four years? Or is it identification creep? You just had more claims piling up and since you were in a one-way situation (once chronic, always chronic), maybe that was identification creep. That may be very well what's happening. The chronic population annualized trend was 5.6 percent. The non-chronic was 13.8 percent, which is very different. You would reject the null hypothesis that the trends should be equal.

We said that maybe what's really happening is that the service sector composition is very different for chronics and non-chronics. In fact, you'd expect it to be different. Chronic people ought to be in the hospital more, for example. Maybe they're taking more drugs. What would the non-chronic trend be if non-chronics had the chronic people's service mix? It didn't change things much. It reduced the non-chronic trend from 13.8 percent to 13.2 percent.

Maybe "once chronic, always chronic" isn't so great. Let's do the "ever/never." That worked a little better. If you actually believe that the trends should be the same, you would say that this must be a good deal here. The chronic trend was now 16.3 percent and the non-chronic trend was 17.2 percent. I didn't do any confidence intervals, but I wouldn't be terribly surprised if they overlapped. Is that the best way to do things? The only problem I have with that is that while I may have had heart failure for a long time before my doctor recognized it, I probably didn't have it forever. So if I had to have a whole bunch of claims before the system would say "you're chronic because you have heart failure," then yes, there was probably some period of time when I had heart failure before a claims-based system would find out that I did, but it probably wasn't a really long time. There would be some clinical questions related to that.

Perhaps trends differ because chronic and non-chronic differ in risk. Now we risk-adjusted using a standard risk-adjuster predictive model using the prospective score (that's where I predict your cost for next year). We adjusted the chronic group, using the prospective risk, to try to get rid of any variability between chronics and non-chronics that was due to their risk (for instance, me just having heart failure or this person over here having seven other things). That came out pretty well. The chronic trend was now 12.5 percent and the non-chronic trend was 11.9 percent. I could believe that that supports the hypothesis that they're equal.

The only problem is that we have an a priori idea that they should be equal. The purpose of this study was not to prove whether they're equal or not. It was to support this general idea that Michael was bringing up earlier that nobody knows whether they're equal or not, but you can probably mess around with them one way or another, and perhaps—this is the subject of another paper that we'll present at the DMAA this fall—what could be done in these reconciliations where vendors have to go to the table with the buyers at the end of the period (hopefully they'll do it at the beginning to avoid a lot of rather predictable headaches that happen at the end) is that they'll say, "What about the prebaseline years?" So now, instead of doing the baseline year and the study year (a trend that's two years long), what about the prebaseline-to-baseline trend? If you could show that the chronics and non-chronics had a prebaseline-to-baseline trend that was equal and then they diverged, while it doesn't prove causality because it's still not experimental (you're still not allocating to separate groups; there's still no nonintervened group), at least it's easier for me to buy the notion, "Before you touched us the trends were equal, and now they're not."

Let's talk about conclusions. Again, how you identify chronic and non-chronic makes a difference. I didn't get into how we identified these people. We used what I would consider to be relatively loose criteria: a couple of face-to-face encounters or a hospitalization with a principal diagnosis of one of those five chronic diseases. Some companies will use much more stringent criteria, such as point-based counting systems and so on, because they don't want to bother a lot of people that don't really have diabetes but just saw the doctor twice and they were ruling it out. I can assure you that using looser criteria makes a big difference. Also, don't assume the non-chronic trend is a valid estimator of chronic trend. It might be helpful, again, to do other studies and see if you can correlate this with one of the other study designs.

When chronics are identified using a prospective "once chronic, always chronic" algorithm, then the unadjusted non-chronic trend is not a good proxy for the chronic trend in DM savings evaluations, because of migration bias. The cheaper people that are just sliding over the line to chronic in a given year sometimes then will dilute out the chronic trend. When using chronic identification algorithms that classify members as "never or always," you get closer, but there are clinical questions about whether that's a good idea. There is one algorithm that we did not

use in this particular study, and that's having to re-qualify every year. I want to be explicit about that because some DM vendors do use that.

Adjusting the chronic trend for service mix has very little effect. We also found, by the way, that eliminating outliers didn't have that much effect. Adjusting for the effect of change in population risk results in an estimated non-chronic trend that *does* closely approximate the chronic trend, which was helpful (using the "once/always" algorithm). When using a prospective "once chronic/always chronic" algorithm, you need to use risk as a risk adjustment in order to make these things useful, at least according to our experiment. Again, you should look at the prebaseline years; make sure that in the "untouched" scenario you are getting trends that are close to each other.

MR. FUHRER: Our third presentation is by Nancy Garrett, who's with HealthPartners in Minneapolis. She has a Ph.D. in demography.

DR. NANCY A. GARRETT: Everyone talks about the aging of the Baby Boomer generation, and there is lots of rhetoric and worry about how we're going to finance care in the future. We wanted to get a little more precise about it in our study. What if we modeled this and looked at not just what happens overall, but also by specific chronic diseases that we know are going to increase in prevalence? What's going to happen in the future? What is the impact of aging specifically going to be on costs? That's what we did in our study.

I'm going to review our methods and data sources for our modeling. We used claims and enrollment data from HealthPartners as a data source. We're a health plan in Minnesota. We used two different years so that we could have 1.2 million members included in our analysis, so that we had a large n. We included all three of our products, which are commercial, Medicaid and Medicare Plus Choice (a Medicare Advantage plan). We looked at costs in terms of the total dollar amount paid to the provider, which is the societal view of cost, including the member liability as well. Again, our objective was to estimate what the impact of aging on total health-care costs would be for the chronic and/or expensive conditions that are of concern to employers over the next five decades.

Our second data source was data from the national data set, the Medical Expenditure Panel Survey (MEPS), in order to take our data and standardize them to a national cost structure. What we have then are not actual HealthPartners costs, but they are standardized costs to put them into a national perspective. Our third data source was U.S. population projections by age and gender in one-year intervals from the census.

We took our claims data and grouped them into episodes of care. We used the Episode Treatment Grouper from Symmetry to do that. Many of you are probably familiar with that, but let me give you a quick overview. It integrates cost and utilization data from multiple sources, and then it groups them together, using

clinical logic, into episodes that represent all services related to treatment for a particular episode of care for an individual. The method is replicable and reputable.

We did some pharmacy data estimation as well, because not all of our members had pharmacy coverage. We really wanted to include pharmacy data as a component of our model, especially given that it is going to be part of Medicare in the future.

Our modeling approach was to apply the estimated U.S. age-, gender- and condition-specific annualized costs, which we estimated from our data, to the projected population in each age and gender group in future years. It's a very straightforward model. We're holding everything constant; we're making some assumptions that our current age- and gender-specific patterns will hold true in the future and applying those to the population as it ages.

What are we going to look like in the future? We're going to be older and bigger. Our population is going to be larger. In our study, we assume that the population size is constant in order to look at just the effect of the change-of-age structure. As we go forward in time, our population becomes older, which is typical of a lot of industrialized countries. In the year 2020, if you look at the top age category, 100+, it starts to get quite a bit bigger, and there are many more females than males, reflecting lower mortality rates for females.

Then we looked at the age- and disease-specific costs per member for these major chronic diseases. Looking at all services, babies are quite expensive. There are higher costs among females during the childbearing ages. There are higher costs for males and females in the older ages. There's an interesting drop-off as well. Often data is summarized into an 85+ age group, and you may miss an interesting but very real drop-off that we've seen in our data as well. In some ways it reflects some of the limits of medical care among those very old ages. If you're 100 years old and have a particular health issue, having surgery to correct it may kill you. There are some limits to what medical interventions can do right now, which is part of why we see that drop-off.

Now I'm going to talk about some specific diseases. Costs are higher among males with diabetes than among females. There's a peak from age 50 to age 55, so it's not just among seniors that you have high costs for diabetes. Regarding CAD, there's quite a large gap between males and females. CHF is very much a disease of the elderly. We don't see many costs for that until we get up into those senior ages. There's an interesting crossover for asthma. We have higher costs among females than males for most of the age groups, except for children and seniors. There are some interesting patterns in psychiatry. For chemical dependency, there's a big spike among ages 15-19, and males are higher than females in all the age groups. Obstetrics, of course, looks a little different.

We took these age-specific costs and applied them to the projected future population to see what the impact of aging will be. For the United States as a whole, we're projecting an 18 percent increase in costs, due to aging alone, by 2050. This is right along the lines of what has been published in the literature; a couple of other models have done this as well. It's interesting to talk to other people about this number. A lot of people are surprised that this number is so low, because that works out to be only a 0.3 percent annual increase. When you're talking about double-digit increases that we're dealing with every year with health care, that seems pretty small to a lot of people.

Then we looked specifically at the chronic diseases that we included. We wanted to see what the change would be and compare that change to the 18 percent average. CHF is affected the most. That has a 75 percent increase as a result of aging. CHF is a disease of the elderly, so that's why it is so impacted by aging, assuming that those patterns would stay the same. CAD had a 48 percent increase. Diabetes had a 24 percent increase, so it is also affected by aging but not quite as much. Some conditions are actually going to decrease as a result of aging. Pregnancy and infertility care have a 12 percent decrease because of that change in age structure. When you hear about the aging of the population, you don't often hear that there is going to be some re-allocation among the different types of services that we do. There is not an increase everywhere. Chemical dependency would have a decrease, also.

Some other numbers give some context around the level of cost. For example, congestive heart failure is going to increase a lot, but it only makes up 1 percent of total costs right now. Even with this 75 percent change, it will still only be 1 percent in 2050. That puts it in some context. Coronary artery disease, on the other hand, is going to increase quite a bit and is also very expensive. It's 7 percent of costs. We might want to be paying more attention to some of those conditions in terms of the impact of aging.

To summarize, the overall change that we saw was 18 percent, about a 0.3 percent annual increase. Most of the increase happens between 2000 and 2030. Behind that average, some of the conditions are going to increase quite a bit more than average, and aging is actually going to reduce costs for some conditions.

Going along with the theme of this conference, we wanted to look at some acute conditions and compare them with the chronic conditions. Based on another study led by Louise Anderson, we looked at a few acute conditions that are categorized based on an episode with a clean window of less than 90 days. They're conditions that are fairly short in duration. In our data, we looked at some of the most frequent acute conditions and some of the most costly acute conditions. These are single episodes, whereas previously we looked at groups of episodes.

For minor visual disturbances, females have higher costs than males for many of the ages. You have some costs in the young ages as well as the old ages; it's not

just elderly people having this particular episode. For hernias, males have higher costs than females. We modeled out these conditions and others (including isolated signs and symptoms, minor inflammation and gall bladder) to see the impact of aging. Hernias had the biggest change, which was 19 percent. That's about average. The others were quite a bit lower; they were more in the single digits. They were not very impacted by aging.

We concluded that most of the acute conditions we looked at were affected less by aging than the chronic conditions, as we would expect and as we've heard other speakers say today. This is some empirical back-up for that. These conditions also make up a small proportion of total costs. Again, that's another reason why they might not be as much of a concern.

MR. FUHRER: Our final presenter is Jaan Sidorov. He's a doctor with Geisinger Health Plan in Danville, Pa.

DR. JAAN SIDOROV: I want to put a special emphasis on what the findings mean to the DM industry, so that you can have an anticipation of what the DM folks may pitch or may be thinking about. There are implications for DM and a more futuristic implication in terms of the Boomers.

In case you don't want to memorize the definition of DM, I know of a cartoon that shows both the good and the ugly as far as the implications of DM. In the cartoon, there's a hapless victim keeled over in a crowded theater. Instead of calling out so see if there's a doctor in the house, the question is, "Is there a case manager in the house?" The upside here, medically, is that what this patient needs is basic life support and then advanced cardiac life support (ACLS). Both basic life support and ACLS are not necessarily in the domain of expertise of physicians. In other words, nurses and other health professionals can help save this individual. Then the patient is going to spend some time in the intensive care unit in the hospital (that's where I come in), and then the patient is going to move out. Now we are going to need the case manager, who hopefully is ACLS-certified. In the beginning, the ACLS manager is going to be necessary to help this patient traverse through this episode of care and deal with benefits, limitations and all the other mechanics of insurance while going through rehabilitation. It's not necessarily such a bad thing to have a case manager helping to lead this patient through the health-care system. That's what DM is all about.

The "ugly" here is that, even though all of you are actuaries and even though you may all be believers in DM, the fact is that you still expect a doctor to give you some face-to-face time during the course of any of your illnesses. But imagine what the doctors in the audience are thinking here in terms of the territoriality of their professional domain and having a bunch of nurses take over this part of the health-care system. That's what I deal with as a medical director advocate of DM. Doctors don't like this scenario very much.

In terms of the incident-versus-prevalent discussions we heard from Dr. Cousins and Dr. Jester, the main knife sticking in the throat of the DM industry nation-wide right now is this whole phenomenon of regression to the mean. It keeps coming up, and it's still a very big issue in helping us credibly ascertain the causality of DM in affecting trends of health-care costs. There was one reference to pre-pre-incident. I will tell you as a medical director that the closest thing in clinical practice of which you may be aware is that there's hypoparathyroidism, pseudohypoparathyroidism and pseudopseudohypoparathyroidism. So if you're talking to doctors about pre-pre- chronic, you may want to also point out to them that this is not that dissimilar from this condition of pseudopseudohypoparathyroidism. The point is that there's a very good fix here. Instead of turning to the randomized control clinical trial, the pre-post analysis, although there's a downside, appears to be a methodology that's getting us closer to this idea that there are mutually exclusive and comparable populations (depending on how well you define them), weighted by member months, that can give you an idea about whether or not DM is affecting trend at all.

We heard a little about DM versus disease diagnosis versus disease claim (just because I diagnose it, there may or may not be a claim for it). The bottom line, as all of us know, is that the Congressional Budget Office and many other policy-makers in Washington, D.C., who are ultimately responsible for financing health care, have come to the conclusion that there's insufficient evidence to conclude that DM programs really do reduce the overall cost of health care, for a lot of different reasons.

There was some discussion about coming up with historical cost methodologies and the use of non-chronic populations to help us better define what's going on out there in terms of trends. The interesting things to me and to the DM industry are better defining the drivers of cost increases, what can mitigate them and, if mitigation is possible, how you design DM programs.

I had a very interesting discussion over lunch today with an actuary who pointed out to me that there's a rising group of actuaries with special expertise and interest in DM. I asked him, "What's the consensus? Do you think it works, or doesn't it?" He said that it depends on the DM program. That's absolutely true. I think that this is another role for actuaries or for those knowledgeable in the industry to understand. This is a DM nurse who's calling up patients to remind them to take their beta blocker because they have CAD, but the fact is that she's inside a very big black box. We don't know to this day exactly what nurses do and how they do it in order to get patients to do the right thing in terms of their health care. There are many different ways that nurses can approach patients. There are many different ways of motivating patients to do the right thing. Understanding those differences can make the difference in terms of the success or failure of a DM program.

As far as the Boomers coming, I thought the data was very interesting from the point of view of the DM industry and what they're thinking about. If we have a population of patients who are elderly and in there, there's a smaller group of

patients who are candidates for DM. If you're thinking about DM as a hammer that can be applied to nails, the question is, what additional nails exist in the population that's chronically getting older, in terms of to what you can apply these interventions? For example, the DM industry is very interested in getting Medicare to pay for DM services. That's what CCIP is all about, and everybody is holding their breath and hoping that that finds a positive result. But in addition to finding additional payers, the hunt is on in the DM industry to find additional diseases to go after, such as obesity or psychiatric illness. Predicting what diseases are going to be out there is very interesting to them.

If you haven't read it, there's a great book on the topic, *Age Wave*, by Ken Dychtwald. He's making quite a living going around and talking about what the Boomers are up to. I think one of the reasons that the DM industry has a future, no matter what, is that people like Jerry Garcia will always welcome additional health care over and beyond what the health-care system is already providing. If Jerry Garcia were alive, he would be able to see me as often as he wanted, and I would try to get him to stop drinking and stop partying and start controlling his diabetes better. But Jerry would always welcome a nurse calling him up and helping him better manage his disease.

MR. IAN DUNCAN: Nancy, I'm very interested in these projections of chronic costs as the population ages. You said that the costs would increase 18 percent due to the aging of the population between now and 2050. To what increase in the average age of the entire population does that correlate? Can we translate that back into so much per year of age increase?
DR. GARRETT: So you were interested in how the average age of the population is changing as we go forward?

MR. DUNCAN: Exactly. If you assumed it was 40 today, what will it be by 2050?

DR. GARRETT: I don't have that information handy, but we do have the data, and it goes up as well.

MR. DUNCAN: If I understood the paper correctly, you're saying that this is a constant population, essentially. The size is constant and the costs have increased by 18 percent, but then the average age has gone up by a certain amount. I'm trying to correlate the two and see how much per year of age the population is increasing. As actuaries, we have a generally accepted theory that costs increase x percent per year of age, and I'm trying to relate it to that.

DR. GARRETT: I don't have that data quite handy.

MR. HOBSON D. CARROLL: Dr. Sidorov was talking about DM being a hammer and looking for nails. I'm reminded of a statement that says that when your only tool is a hammer, you have a tendency to perceive all problems as a nail.

MR. FUHRER: Nancy, were your age tables based on raw estimates or was any age-to-age smoothing done?

DR. GARRETT: Those were actual estimates. The only thing we did to change them was to standardize them to the national data. They're in five-year age groups. We got the population data from the census, but the health-care data is from our health plan.