RECORD, Volume 29, No. 2

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

Session 21PD
New Strategies in Disease and Utilization Management: Substituting Facts for Assumptions

Track: Health

Moderator: IAN G. DUNCAN
Panelists: IAN G. DUNCAN
ALISON JOHNSON†
THOMAS WILSON‡

Summary: Disease Management (DM) programs are becoming common among health insurers. Some are convinced that these programs are of significant value in controlling health-care utilization, while others are not so sure. Actuaries are at risk of losing responsibility for the construction and evaluation of such programs to economists and biostatisticians. This session features members of the DM industry and health actuaries who work with them.

MR. IAN G. DUNCAN: Care management implies some activity on the part of the care manager—whether it's a health plan provider, case manager or another professional responsible for the care of the individual. Has the activity of the care manager or the intervention added value in terms of reducing resource utilization for a constant level of well-being or increasing well-being for a constant level of resource allocation? Different care managers, all of whom seem to have a financial stake in the outcome, make claims about their value. So actuaries and those responsible for the financial management of health plans are often called upon to

†Ms. Johnson, not a member of the sponsoring organizations, is a Consultant at Milliman USA in Minneapolis, Minn.
‡Mr. Wilson, not a member of the sponsoring organizations, is an epidemiologist at Wilson Research, LLC in Loveland, Ohio.
Note: The chart(s) referred to in the text can be found at the end of the manuscript.

advise health plan management on financial and related issues concerning interventions.
1. Did the intervention have an effect, and can that effect be quantified?

2. Should the health plan perform more or less of a specific intervention or perhaps change the intervention target?

3. Are there any other potential interventions that might have more value for a health plan that the health plan could consider applying?

How should the actuary go about answering these questions? The standard actuarial tools do not give us a whole lot of help. Many health plan medical management departments seem to be turning to health economists and sources other than actuaries to answer these questions. And I think that with the right tools and techniques, actuaries can make a significant contribution in this area. For that reason, we have gone outside of the actuarial profession this afternoon and invited two speakers with specific expertise.

Disease management is a growing area for managed care companies and plan sponsors. Despite significant fees that companies charge for their programs, the financial potential of the programs is quite significant. The controversy over their value is significant, as well. Fortunately, the opportunities for actuaries seem to be changing. We are called on more often to offer our expertise.

I want to draw your attention to three recent publications that are relevant in this area. The first is called "Standard Outcome Metrics and a Valuation Methodology for Disease Management Programs." It was published by American Healthways and Johns Hopkins University. The American Healthways paper notes that the vast majority of programs operated by health plans are run as part of their core business, not as part of a research effort. The authors of the paper try to find ways to measure outcomes in a reasonably uncontroversial way in a situation in which you cannot run a controlled experiment. I think that the American Healthways paper is quite useful because it contains numerous clinical measures that are worth paying attention to in this area. Their concentration on financial measures and outcomes is less, I think, than their clinical focus. For financial measurement, they seem to favor a method that I would call adjusted historical control (i.e., the experience of a similarly defined population in the year prior to program inception is used as a baseline and projected forward with trend to the measurement period, and then compared with actual experience).

The second paper that I draw to your attention is by one of today's speakers, Thomas Wilson. It's called, "Strategies for Assessing Causality in Disease Management Programs." It was commissioned by the Disease Management Association of America and will be published in the journal, *Disease Management*.

And the third is an article called, "A Prediction Model for Targeting Low-Cost, High-Risk Members of Managed Care Organizations," which I co-wrote with colleagues. It is not about outcomes measurement, but rather it is about the use of prediction to support the economic aspects of disease management programs.

We will have three speakers on our panel this afternoon—I referred to Mr. Wilson already. He's an epidemiologist, a frequent speaker, and one of the authors of the Disease Management Association of America's paper on outcomes.

Alison Johnson from Milliman USA, also will be presenting. Ms. Johnson is the author of a recent study of the disease management industry published by Milliman USA, entitled, "Disease Management: The Programs and the Promise." She's a registered nurse and holds an MBA degree. Her career includes clinical positions as well as health plan positions. She is also the author of a chapter in the recently published fifth edition of a book that we are all intimately familiar with, "Group Insurance," edited by Bill Bluhm. That chapter is called "Medical Care Management."

I am also going to speak. I am a consulting actuary with Lotter Actuarial Partners, Inc. My practice focuses on disease management, outcomes measurement and achievement of financial results. For two years I headed a small disease management company at which I implemented and managed programs that served approximately 100,000 chronically ill members of several health plans.

I entitled my presentation, "Understanding the Economics of Disease Management." Mr. Wilson is going to talk about epidemiological strategies. Ms. Johnson is going to present some of the results from the Milliman USA study. If there is time at the end, I will present some data that we have been analyzing from programs that we ran, which include control-group data.

Chart 1 shows the environment in which we find ourselves. The days of fairly low trend are over. Trend is escalating again. There is a general view that cost increases are driven by a small number of high utilizers. And some health plans say that "X" percent of our population drives "Y" percent of our cost, as though that somehow resolves the problem. In fact, it does not. And it's not a particularly useful statement. What is important is finding future high-cost people, those who will form part of next year's "X" percent, and who you have some chance of managing toward a reduced financial outcome. And since cost management implies that something did not happen, you have the added problem of trying to prove, through data, a negative (what did not happen).

If we could find the right people and manage them, we could reduce our cost. Who are the right people? How do we manage them? What kind of programs can we use to reduce costs? What amount of cost reduction can realistically be expected from different groups of people? And how do we measure the cost reduction?

People's costs in health plans are dynamic. Individual member costs are constantly either rising or falling, if left on their own.

That's both a blessing and a curse. It is a blessing, because if you can find those people whose costs are on the increase, then you might be able to intervene and reduce their costs. On the other hand, it's a curse, because the fact that costs are falling absent any kind of intervention gives disease management companies the opportunity to claim that reduction as a result of their intervention and pocket the fee. But some of these numbers will show you that there is, within any population, a significant percentage of members who, in the course of a year, move from low cost to high cost, as well as from high cost to low cost. And part of the trick is identifying those people in advance and intervening with them.

Chart 2 draws your attention to the work of Dr. Jack Wennberg. For those of you who are not familiar with him, Dr. Wennberg's contribution to science has been the ideas of unwarranted variation in utilization of medical services, and "Supply-induced demand." One of the things that he's pointed out to us is the percentage of health-care costs that are optional, or arise from causes that in fact are manageable or intervenable.

If you add up the different bars in this chart, about 20 percent of the population is going to account for 50 percent of the costs that are also manageable. The studies of Dr. Wennberg and his colleagues have shown that if you intervene with people who are going to have procedures or incur costs for certain diseases or procedures, giving people the right kind of information at a time when they're receptive to this information can result in the patient choosing alternative, less-invasive and less-costly procedures, leading to cost reduction.

There is a basic concept that we call the yield curve. This is a continuous ranking of people according to one particular statistic—the probability that they will have whatever the predicted event is. And a predicted event could be, for example, readmission to a hospital. Take a group of people who were discharged from a hospital—in a Medicare population the frequency of readmission within 90 days could be somewhere between 25 and 35 percent. If you want to intervene with all of those people, it would be costly. Choosing people for intervention at random, the chances of making a difference and having a program that is cost effective would be relatively low. On the other hand, if you can "risk rank" them, and then intervene quickest with the people who have the highest probability of re-admission, you will be able to produce a better financial outcome.

A lot of the work that's done in predictive modeling deals with predicting cost. The approach that we took in our article was to show that by holding costs constant or holding the event constant—like a hospital admission—you could focus on these frequencies. If you rank members according to their frequency of expected event, that is then useful for managing a program. You can further convert this yield curve into a financial-opportunity curve by including variables like the opportunity

to reach and engage individuals, the cost of their hospitalization and the likelihood that once you reach them, you can actually change their behavior.

Now there are some key drivers of the economic model that are important to bear in mind. This is not all about prediction. One of them is the prevalence of whatever it is that you are trying to manage within the population. High prevalence of high-cost conditions, such as congestive heart failure, will generate a financial opportunity. If the high-risk members are not there (low prevalence) the financial outcomes will not be, either. One factor influencing financial outcome that I stress frequently with my clients is data quality. A great deal of focus has been placed on prediction and the use of prediction to identify high-risk patients within health plans. On the other hand, sometimes health plans have bad data on, for example, the telephone numbers of people that they are trying to reach. If they only have quality telephone numbers on 75 percent of their members, it really doesn't matter how good your prediction algorithms are, since at best, you're only going to be able to reach 75 percent of the population.

In my opinion, some of the work that needs to be done in the economics of these programs involves determining the sensitivity of the financial outcome that you are seeking to the key variables. Where does it make sense to put our investment? In my previous example, improving the prediction algorithms may be less financially beneficial than simply getting telephone numbers on 100 percent of the population. The other variables include the ability to reach and engage members, the ability to actually change people's behavior through intervention and the productivity of your resources. Clinical resources are expensive. It is very important that these resources be applied where they can do the most good in a timely fashion.

Chart 3 brings all of the financial aspects together in a very simple graph. As you intervene with a population, you combine all of these variables: reach and engage rate, the probability of events, the probably to change people's behavior, etc. You can combine the data into a gross savings curve. Gross savings increases at first quite rapidly as you penetrate the highest-risk population. But as the probability of that event starts to decline, the increase in your gross savings "tails off" quite quickly. On the other hand, as you reach down further into the population, you're managing larger and larger numbers of people, so the cost of the interventions rises. Obviously, your net savings (the difference between the two) falls rapidly and becomes negative. How broad a program and how deeply into a population you reach depends on the relative slope of these two lines.

What I do not have on Chart 3 is return on investment. This is something that you hear about all the time. Vendors are asked the question, "What is the ROI on your program?" as though that question has a unique answer. There is a different "return on investment" number at each point on the graph; ROI starts off very high, declines towards one and becomes less than one at the point that the gross savings and cost lines cross. I hope that you've seen enough of this economic approach to understand that there really isn't a unique answer to the question:

"What is the ROI of your program?" In addition to all the key variables mentioned above, it depends on where you want to draw the intervention line. ROI, the relationship between savings and cost may not be the right number to focus on, anyway. The right number to focus on might be absolute savings, because you can have a fabulous ROI on a small number of people and have absolutely no noticeable effect on the health plan's bottom line.

Finally, to round out the economic discussion, economic outcomes from these kinds of programs are very widely distributed around the mean, in my experience. Chart 4 shows real data, slightly modified, from one of our programs. I have related the quarterly savings on a per-month per-member (PMPM) basis from an intervened population, measuring it over different quarters. It is a very highly skewed distribution, with a long tail to the right. It shows you two interesting things. You could actually have a very low measured-savings outcome at a particular point in time (or even negative savings) and still be within the 95th percentile of the mean. That is a nice observation for DM companies, because it allows a DM Vendor, to go to a health plan and say, "We're measuring negative savings, but it is within the 95th percentile, so really the answer's positive; you cannot reject the hypothesis that savings are at the predicted level." I think that it is important that people; when they put these kinds of programs in place, realize that the results are going to fluctuate very wildly, and that an apparently low savings outcome may not represent failure.

And the second observation from this data is that, if you understand the distribution and the likelihood of these outcomes over time, you can expect some form of guarantee from a disease management company regarding their savings. The price for the guarantee, as you look at those distributions, might be quite high. But at least, if you understand where the outcome is likely to be on a distribution, you can think in terms of getting that guarantee.

So, that's my brief look at the economics of disease management. I hope that I have convinced you that there is no single answer to the question, "What's the ROI of your program?"

MR. THOMAS WILSON: I'm going to talk about epidemiology strategies for disease management, substituting facts for assumptions. I used to work at Anthem Blue Cross/Blue Shield as their corporate epidemiologist. I didn't do any data analysis there, but when I went out on my own in 1999, I realized that all of the things that I wanted to do and needed to do I couldn't do, because of the way the data was organized. I actually went back and created a whole new system, so I could do the work that one needs to do as an epidemiologist. A lot of the analysis that I have done is based upon that system.

I am going to talk about pragmatic epidemiology, as distinguished from epidemiology. I call it the "epidemiology of value." It is not just looking at disease states and distributions from an academic, objective standpoint. What we really

want to do is change things. So it is an action-oriented science. I define it this way: It is the scientific study of the distribution and determinants of health-related value in defined populations and the application of this study to the control of health-related value problems. This is really a classic definition of epidemiology, except for the word "value." Most epidemiologists are only concerned about health and disease. In this industry, we need to think about other forms of value, financial and otherwise.

I see value in three different spheres, and it is really dependent on where you are in the company and what kind of value you are interested in. There is a value that I call health—people's health, clinical issues. There's a value called economics—financial, return on investment. And there is another kind of value that I call perception or satisfaction. And depending on where you are in these firms, you want one or the other. If you're in the marketing department, you want to increase satisfaction. If you're the chief financial officer, you want a return on investment. If you are the chief medical officer, you want people's blood pressure to go down, or whatever. And there are fights over these things. I think that the outcome that you are choosing to predict and build probability models on is never objective. It is always subjective. That is a really important point to make. I think a lot of the arguments that we have are caused by the different outcomes that we are interested in.

If, in fact, we had value, we want to know what caused that value. Was it something that we did? Something that the marketing department did? Or would it have happened anyway? So the big question is, what would have happened to the disease management (DM) population in the absence of a disease manager and intervention? We can never know the answer to that question, because people cannot be in two places at the same time. This particular question is why epidemiology exists as a discipline and has been around for at least 150 years. We want to know what would have happened had something not occurred, had this germ not been there, had sewers not been built, etc.

The only way to answer that question is to have a comparison group or a reference group. Chart 5 shows how you would measure impact—and this could be clinical impact, satisfaction impact, financial impact. Essentially, it is easy. You have to find out the pattern of whatever you're measuring in a reference population. In other words, what would have happened to that population had they not had the intervention? In this example, it is the darker line. Once you know that, everything is easy. All you have to do is measure what actually did happen, and then you compare what actually did happen to what you expected would happen. The difference between those two lines is a measure of impact, the impact of your program. Now this only works if these two populations are equivalent to each other.

Where does the "expected line" come from? If you want to answer one question, that is the one you have to answer. There's too much of a focus on trying to get

better metrics and better data when the real issue is, where does the expected come from? How valid is your measure of the expected?

I mentioned the term "equivalent." A lot of people believe that if you do a randomized clinical trial, you have done everything. The only reason that occurs in the best of circumstances is the people you randomized into the intervention group were equivalent to those randomized in the control group. Randomization is a tool by which you can get equivalence. There is no guarantee that you are going to get it, but it is a great tool to try to get it. And often, you do not get it on every important variable. So you have to use adjustment techniques.

Once you have two equivalent populations—one that has what you are interested in, and one that does not—you have to measure things. And there are really three kinds of things you measure—what we call Type One metrics, Type Two metrics and Type Three metrics.

Type One metrics are the easiest things to measure. Those are the things that you do in your intervention. You call people on the phone. You give people a blood-pressure test. You send out a postcard. Those are really easy to measure. They are overemphasized in their importance, but it is good to know that the company that you hired is actually doing something. This is a way of measuring what it does.

The Type Two metric is a much more important metric. It says that you did something (a Type One metric). What happened as a result of what you did? What do you think will happen? "I think that the blood pressure will go down." These are never known for sure.

Cost is actually a Type Three metric, because in disease management, the goal is to improve health. By improving health, we decrease cost. So the improvement of health is a Type Two metric. The decrease in cost is a Type Three metric. It's just a standard hypothesis or a causal pathway. Type One activities and metrics represented by them cause Type Two metrics, which cause Type Three metrics. It's a cause/effect relationship.

Comparability—people get confused by this. Basically, you want to know if the metrics that you measured in the reference group were measured in the same way in the intervention group? Because if you are only using claims data, it is probably true that they are comparable. But in health care, it often is not true, because people are basing measured impact on peer-reviewed journal articles. And they say, "In this study, this improvement occurred. Therefore, we did the same thing in our study, and we are going to say that improvement occurred." Well, were things measured in the same way? Were the inclusion criteria the same? The exclusion criteria? Was blood pressure measured the same? You need comparable metrics.

Ultimately, to do a good evaluation of return on investment or any kind of impact, you have to have two populations that are equivalent to each other and three types of metrics that are comparable to each other.

How do you get this information? There are multiple ways to try to get the "expected line." In epidemiology, we talk about four kinds of studies. In our paper, we talk about seven different kinds. I am only going to talk about three of them. One I call a post-only design, another is a quasi-experimental design and the last one is a follow-up or "cohort" design.

A post-only design measures the people that have the intervention—like in a case-management program. What was their baseline metric? A year later, you measure them again, and you get some answer. You can do that at a population level. There are a lot of problems with this, one of which we call "regression to the mean." Chart 6 shows an example. I pick a certain cost threshold in a two-month period, and I ask, "What percentage of the people are above that cost threshold?" And then the next month, what percentage of the people are above the cost threshold?

In the first month, I picked everybody that was above the cost threshold. So 100 percent of the people, by definition, are above the cost threshold in the first month. In the next month, what percentage of those people are above the cost threshold? Well, through pure random error, 1 percent, because I picked the 99th percentile as my cost threshold. That's regression to the mean.

Chart 6 shows results for actual diseases. Basically, if regression to the mean was 100 percent operative, you would have this relationship between 100 percent and 1 percent in the first part. But, in fact, you do not. If I look at otitis media, and in the second month 14 percent are still above the threshold, that means that from a clinical standpoint and an insurance standpoint, there is an opportunity. It did not go down to 1 percent. There's still a need for medical care in this country. But I think that people are overestimating, throughout history, the impact of medical care on high-risk people—which, of course, is what medicine is all about, treating high-risk people. There are six or seven different diseases for which you see that regression to the mean varies by disease.

A second type of study is a quasi-experimental study. It is a pre/post study for which the patients are their own control. And although they say that there is no reference population, the pre-period is the reference population for the post-period. And the separation between the two is when you start your disease management program. Essentially, you're going to use the period before the intervention as a surrogate, as a proxy for what would have happened in the absence of the DM intervention in the post-period.

This is a great design. I am looking at the pre-period. This is just a measure of sickness. And things are not constant in health, they fluctuate. A population fluctuates just like an individual fluctuates. So if you have the kind of fluctuation as

in the pre-period and post-period figures in Chart 7, the pre-period is a pretty good indication of the post-period. So any change that might occur could be due to your intervention.

This is where regression to the mean comes in. It depends on when you measure people in the cycle, because there are cycles, as shown in Chart 8. If you're a person that was measured at the top of the cycle in the pre-period, at the bottom of the cycle in the post-period, and you have a lot of that being done, it'll appear that things actually improved when, in fact, it's totally spurious. Alternatively, you could measure people on the left side at the low end of the cycle and measure them at the right side at the high end of the cycle, and you would show that things actually progressed, which also is spurious. So it is a classic regression to the mean when you say, "Let's only pick the people in the pre-period that are really high-risk." When you do that, just by random chance, you will likely have an improvement. So that's regression to the mean. But there are a lot of other problems with this kind of design.

If you have the kind of pattern in Chart 9, for which things are actually getting worse, let's call it sickness, (high is bad and low is good) it fluctuates, but it is fluctuating upward. And the moment that you find these people, it actually starts going down again. In that situation, a pre/post design is awful, because you are going to project the pre-period. And if you do it like a linear model or something, you are going to project up, and it actually goes down in the absence of any intervention, and you do not know it. The disease management company is going to get a lot of credit for something that they did not do.

Chart 10 shows examples for which the pre/post would be a good idea, if you were aware of these patterns. A lot of people in health care talk about necrotic progressive disease. It's getting worse over time. The top of Chart 10 shows that things are progressively getting worse. If you know that, you can adjust for it, and a pre/post design is OK. On the bottom of Chart 10, things are actually getting better. If you don't know that, you're going to show this decline and take credit for something that you didn't do.

Another problem is what sort of statistic you're using. Chart 11 shows a real-life example of a congestive heart failure program for which I looked at the average cost before and the average cost after. On the left, the cost went up in this program. They did not like that. So we measured the median, which is the right side, and it went down. That is the one they wanted. So you have to be careful.

Time-series analysis is basically a pre/post design with multiple measurement points. Chart 12 shows another real-life example. We are looking at the incident event of a specific disease—in this case, congestive heart failure. And I am showing what proportion of the people in the incident time segment are above a certain cost threshold. In that incident time segment, a lot of people are above it. They are probably in the hospital. So we are showing that 16 percent of the people are above

the threshold, that is right in the middle at the mountaintop. Prior to that, you see people growing up to that incident event. Afterward, they are getting better on the average—not everybody, but the population average gets better.

Because of the claims lag in our business, we do not find high-risk people until three months—probably later—after it happened. It is sort of like an explosion in outer space. We see it today, but it happened 10 million years ago. We start intervening, at best, in the third month. Now we take the average of the pre-period as our pre-period. The PMPM is going to be really high, and we compare it to the post-period. Already, I can guarantee you that we have won without doing anything.

Chart 13 shows a classic way that you would measure return on investment in a time-series analysis with a control group, a reference population that is equivalent. And there are many valid ways of getting that information without randomization. I'm not going to get into how you do that now, but again, the difference between these two lines is a measure of impact.

So I talked about the epidemiology value, for which we expand epidemiology to look not just at health issues but financial and satisfaction issues. We talked about equivalent, meaning you have got to have a reference group. Even if you don't think you have one, you have one. You have to compare yourself to something. Comparability—metrics have to be measured in the same way in both populations. And by the way, the models that you use, statistical models, have to be done in the same way in both populations, which is another big error. We talked about pre/post designs and a lot of the problems that they have. We talked about some follow-up designs. I mentioned observational and experimental designs. I want to end with one thing. An experimental follow-up design is a randomized control trial, what the FDA does, and those are not without problems.

MS. ALISON JOHNSON: I am going to talk about a research report that Milliman recently released (that I am pleased to say I am the author of). So far, we have been talking about the outcomes of disease management programs. And I know that for actuaries, that is a prime concern. How do you measure the outcome of these programs? I am here to terrify you with some operational concerns about disease management programs. We will turn away from outcomes measurement and talk about the interventions and the operations of them.

This is information from my research report. There are a couple of areas that I am going to give you more information about. First of all, there is information in my report about the conceptual approach. I identified about 115 disease management vendors in the United States, and of those 115, a representative sample of 14 agreed to cooperate and complete a rather lengthy survey, delving into what they do, why they do it and how they do it. I outlined what the conceptual approach is, what various people believe about doing disease management and what the underpinnings of it are. There are some interesting things in the report regarding

the natural progression of chronic disease. How hopeless is chronic disease? How much improvement is possible? Those kinds of things are in the report. Also, the report gives some sense of who controls chronic disease. Individuals with chronic diseases are the ones that are supposed to be in control of it.
I am going to talk about member management quite a bit and will present some of the tables from member management. This is one of those key things that disease management companies really have been struggling with. How do you identify the people that will benefit from disease management programs and then get them enrolled and get them to cooperate with intervention? The report gives clinical information sources, also.

This is something that many clinicians are quiet about. When people refer to the art of medicine, there is a reason for that. If you look at the things that are done to diagnose and treat diseases, the general public has an assumption that those clinicians know what they are doing. They assume that there is a research base that definitively says that this diagnostic study will tell us whether or not they have a disease. This medication will produce this kind of a result.

Does anybody know the percentage of medical practice that uses randomized control trials? It is less than 20 percent. I included information about clinical-information sources in my report. It is not just a matter of who your expert is, but it is a matter of what kind of studies you are looking at.

The report gives a listing of the interventions, because there's quite an array of interventions out there—everything from telephone calls to mailings to educational pieces to assessments of people to actual delivery of durable medical equipment to home visits. I will talk about program effectiveness, too. While I don't have the details of how to measure the effectiveness, I can tell you what companies are doing right now regarding the ROIs that they're reporting and what methods they're currently using to measure their effectiveness.

I am going to talk to you about key findings from various other areas of the report. And then, being an opinionated person, I have a short list of suggested improvements. The back-end of the report actually gives information about the structure of disease management companies. It answers questions about things like staffing ratios. How many nurses do they have for how many people? How many years have these companies been in business? Who owns them? How many lives do they typically cover? Do they operate in conjunction with other medical management services? What about accreditations? How are they insured?

Insurance is an area in which actuaries are concerned about the cost of malpractice insurance and what impact that has. I collected that survey information. Lastly, the report has quite a large annotated bibliography, but the bibliography is from some of those early, very optimistic reports that we saw.

There has been this change in viewpoint in the several years that I have been working with medical management—in particular, with disease management. About a year ago, there were no sales contracts. People were trying to get everybody signed up. It looked like the one piece of medical management that really had some hope of working. There was a focus on developing new contracts, working out all of the nuances in contracting. In the past year, we have seen failed contracts. We have seen losses. We have seen disease management companies that have not been able to deliver on financial promises that they have made. And there is a real focus on fixing the situation right now—including bringing in actuaries to fix problem situations.

Some of the very early and really optimistic reports that came out were everywhere in the medical literature, "A 25 percent reduction in hospitalization for people with congestive heart failure, $50 PMPM for a diabetic." It was just remarkable. They were small studies, and they took a pre/post look at things, or else they looked at current cost and utilization, compared to traditional cost and utilization. It looked like this was the best thing on earth. Health plans were ready to jump on board, because it really looked like it made a difference financially. It also had huge appeal to clinicians, because when you talk about reducing hospitalizations, you are talking about a big quality measure. If you can keep people with chronic illness in better control of their own illness, you undeniably get better-quality outcomes. It also is very appealing to members, in the sense that there is a difference between what the doctor is worried about and what the patient is worried about. The doctor is concerned about biology. Did the lab value improve? And the patient is worried about his or her life. Can I go to work? Can I function? They are close, but they are not identical. And disease management seems to provide both.

Many of those studies were very small. Those early studies would report a handful of patients—30 to 60 patients, maybe. None of them were randomized control trials, and there was no meta-analysis—nobody gathered all of the studies together to do that. And there were no long-term results, yet. They were just looking at as much as a quarter's worth of data. In the more long-term studies, they might look at two years' worth of data—pre and then post. We have some early indications now that regression to the mean and selection bias can be very significant in these disease management programs. That is probably obvious to any actuary that you ask about it, but not obvious to clinicians at all.

Let me talk about finding and enrolling members, because there are basically four different ways that disease management companies are telling us that they find people. When you think about it, claims analysis is not enough. All 14 of the companies that participated in my report did some sort of claims analysis in order to identify members, but there are some problems with claims analysis. One is the lag between when the service was delivered and when the claim actually makes it through the system, gets paid and can be counted. Then there are other issues. So you have just signed up for insurance, and you do not have a claims history that can be mined for that kind of information. What do you do now, wait until you get

sick in order to find out? I see significant problems with claims analysis in that they do not always capture International Classification of Diseases—9th Revision—Clinical Modification (ICD-9-CM) codes. So you will have procedure codes. You will know what was done, but you will not know the underlying diagnosis. And, of course, the underlying diagnosis is key for disease management. So there are many problems with claims analysis, but every disease management company, without a doubt, does it.

The other thing that all disease management companies do is use referral sources. The doctor knows that this patient has got a chronic disease. They've been in his office many times, and they've been hospitalized three times. They're in the emergency room once a month. This is a risky individual. Let's get him over into disease management. Those referrals can come from everywhere. The patient can call in themselves. You can call your mother in. One of the problems is that the number of patients that actually get called in is very small. It is not a priority to call those people into a disease management program. Imagine that you are in the middle of a busy clinic. You have additional work to figure out who their insurer is and what their disease management program is, and find a telephone number, and it is easier to ignore the referral.

It is probably an unusually risky group of members that gets referred in. Generally, the physician will try to manage the disease himself. He tries offering the patient more education, more help. By the time they are to the point where it is worth it for the physician to look up the phone number and call that person into a disease management program, they are fairly frustrated. And this may be a patient who is either wildly complex or very noncompliant. You are getting an unrepresentative group that would be the most unlikely to respond to a disease management approach.

Health-risk-assessment forms are a growing way of collecting information for referring people into disease management programs. Health-risk-assessment forms ask people a variety of questions about their health. It is self-reported health information. The really good thing about health-risk-assessment forms is that you can ask people questions that you never could find on claims. For example, you can ask them if they have limited vision. And that might never show up on a claim. You can ask people if they live alone. You can ask them if they have behavioral health issues. You can ask all kinds of information that really can impact how well you can manage yourself but would never show up on a claim. So it's great from that standpoint.

The real downfall to HRAs is getting the forms back. It looks like the Medicare group is pretty good about returning HRAs, with reported return rates as high as 80 percent. But the commercial population just does not bother with returning them. There are some programs that have a nurse call and ask those questions on the phone. And there are questions about whether or not people really want to have that information known, and whether or not they'll answer questions truthfully.

There are all kinds of issues, but health-risk appraisals actually look like a good way to collect information and identify people for disease management programs, primarily because you can collect social information.

Lastly, there is predictive modeling. So far, we know that predictive modeling does not have great predictive power.

There are some other methods that are used for identifying people for these programs, and let me just list some of them. Instead of looking for a certain claims profile, some people will look for a certain claims trigger. For example, you may have a claims trigger that says that if this ICD-9-CM code shows up, the patient goes into disease management, or something like that. Some programs operate a nurse advice line. People will call in for a fairly innocuous question, and the nurse, in the process of asking questions, recognizes a person that could benefit from disease management, and then he feeds those people into the program.

Some programs have gone to self-enrollment. They will send out all kinds of marketing materials to people, explaining the terrific things they can do for them, and then count on people to call in and sign themselves up. That's another method—member questionnaires, outgoing phone calls to people. Workers' compensation data is an interesting place to find out whether or not people in a workers' compensation situation really have chronic illnesses and are appropriate for programs. And some conditions, such as low back pain, are appropriate for disease management. However, significant diseases like diabetes and congestive heart failure are not typically worker's compensation conditions.

The same is true of disability data. Pharmacy data analysis is another great way to identify candidates, because so many chronic illnesses require taking medication constantly. Mining pharmacy data looks like a great way to examine it. But so much of the claims data and the pharmacy data are not merged, to say nothing of the behavioral-health data that is not merged, making it difficult to identify people in all three of those databases.

Lastly, pre-admission certification—although this sounds like a great method, it actually is not very good, because somewhere in the neighborhood of 60 to 80 percent of hospital admissions are not pre-planned. They are emergencies. They come in through the emergency room. So prior-authorization programs only get at that small subset of premeditated surgeries. All the emergencies, all the medical admissions—none of that is prior-authorized. It sounds good, but it actually does not turn out to be particularly helpful.

Once you identify members, what do you do? You have to get them into the program and find some way to stratify them. Let's talk about targeting members. Disease management companies use several different enrollment methods. They call them or mail them. Four of the 14 programs call *and* mail. All of the companies made multiple phone calls. Companies would make more than one phone call, but

they would not do more than one mailing to people. Some of the problems that are involved are erroneous phone numbers and addresses.

The biggest debate about enrolling people seems to be opt-in versus opt-out. Under an opt-out model, you are in, unless you tell the DM company that you want out. Under an opt-in mode, you are out, unless you tell them that you want in. And when you are evaluating disease management programs, the ones that report really high participation rates are the ones that use the opt-out model. The ones that report participation rates of 30 to 50 percent are the ones that said that you are in, only if you sign up. So I would not even say that one method is better than the other. You just need to know which one you're looking at when you're evaluating participation rates and whether or not that number is reasonable for that model. Most programs talk about going to an opt-in program. They just send you a letter that says, "We've identified from your claims data, or from your risk-assessment form, that you're appropriate for this program, so we put you on the mailing list. You're going to be getting phone calls from us." Unless the member (when the DM company finally gets in touch with him or her) says, "Are you kidding? I don't want anything to do with this."

Once they have got people enrolled, and I have seen a couple of situations for which enrollment was the key to the failure of the program, the company could not swiftly identify and enroll members. After six months, they still had not been able to get the bulk of the people enrolled. So this is a critical piece that the vendor has to do—get people enrolled.

Let us talk about stratification, because this is the next key step. So you found them. You have enrolled them. Next, you have to stratify them. I found that in stratification, people tend to either say that there are two levels, three levels, four levels or they say that every person is an individual and deserving of their own care plan. There is good reason to stratify people. One of the main reasons to stratify people is that instead of having to use a case-management approach—which is one nurse, one patient, one individualized care plan—you can use a more generic approach. Such an approach might say, "Most people with congestive heart failure in Stage One, Class One congestive heart failure, respond to the following interventions." That is what they apply for everybody that they classify into Stage 1, instead of developing an individualized care plan. It is just a lot more efficient if you can put people into strata.

Also, this approach lets the disease management company develop interventions for each different stratification level—your interventions for a Level One, your interventions for a Level Two, etc. And they can make sure that the cost benefit starts to work out. For example, some of the most sophisticated congestive heart failure programs install a scale in the patient's home. Every morning the patient stands on that scale, and the person's weight data is communicated to the nurse electronically. So she knows how much you weigh now, and she then compares that to your previous weight. If you have gained more then two pounds overnight or if

you did not weigh yourself, you get a phone call from the nurse immediately. If you have gained more than two pounds, a medication adjustment occurs. She checks the rest of your symptoms. But you cannot afford to install the scale for all patients. You only want to do it when people need to have medications adjusted constantly.

That's the other thing stratification does, it lets you target the interventions, especially the more expensive interventions, at the people that will benefit from them. You do not want to spend your time chasing down people who would be weighing themselves every day when it really does not make any difference. The program starts to fall apart if you do that. That is what stratification is all about. It fits between one-size-fits-all and a very individualized approach. I think that stratification is a good idea, and it looks as though four levels are about right.

One other word about stratification, some people put patients in Level 1 if they just have risk factors. If they do not have chronic illness but they have risk factors, the patient could be in Level One. And that means that the intervention program targets the patient for some education. For example, smokers might be targeted in a chronic obstructive-pulmonary disease program. Even though they do not have chronic obstructive-pulmonary disease, they are at high risk to develop it. So patients might find themselves stratified into disease management if they smoke, even though they have no active chronic disease.

Calculating cost savings—this is not the actuarial take on what should happen; this is what they're actually doing. In clinical practice, there is a general acceptance that you need to measure four different things any time you are looking at program effectiveness. Cost and utilization are considered together as one of those factors. Another factor is clinical outcomes. Did your lab values improve? Does your liver function better? Does your carotid-artery blood pressure look better? Functional status—are you able to get to work? Are you less depressed than you were? Can you function in life? And then satisfaction—are you happy with the program and with the way that things worked out. Clinicians are constantly looking at those four things every time they evaluate programs. I asked people to tell me about cost and utilization savings, so that's what I'll be talking to you about.

Once you sort through all the verbiage, it basically came down to three different methods that they use. Two of them look OK. And one of them, I think, is downright scary. The first method was to compare individual patient pre-enrollment medical expenses to the same patient's post-enrollment expenses. Sometimes they did that with an adjustment, so that they could make that pre and post measure up. I did not come across a single description that rigorously described regression to the mean, selection bias and the impact that those things had. I did not see anybody that was accounting for it yet, but there was recognition that you had to do something to even up the pre and post.

Method two was to compare the disease management-intervention group to another group that had the disease but was not on the program. The typical way that an insurer would do that would be to offer disease management services as a part of the vendor package to some people, but not to all people. Sometimes the differences between the two were a lot more than just whether or not that service was offered. It was not a great method, although we did see a few people using it.

Method three is one that will scare you. Traditional utilization management is offered by the vendor company and they have decided to add disease management services. Savings are calculated as "days denied," that is, for example, the doctor calls in and asks for 20 physical therapy visits for the patient, but our guidelines say to grant six visits. So we're going to grant six visits, and we're going to count the savings as the other 14 visits. So 14 times the cost of a physical therapy visit is what we save. Now, this method produces some nice numbers. But when you look at the bottom line on the insurance company side, you have got to wonder where the savings are. It does not show up in the per member/per month cost.

Four companies in my study were using method one, and another four companies were using method one with some adjustment. There were two companies that were using method two. And then there were two companies using method three and counting up services that they did not use. There were another two companies that didn't reveal what they were doing.

Self-reported ROI from the companies—I asked people to tell me the cost of administering the program, ignoring any additional cost to the health plan or the employer. If the health plan, typically, has to assign somebody to manage the vendor relationship or if there needs to be an on-site person from the employer managing the program, that is outside of the survey data. So this is just the cost that they pay to the disease management vendor. This amount is compared with the dollars saved.

So here is what people are reporting. The first amount is $5to $1. So for every $1 spent, $5 was saved. The others were $2.90 to $1, $8.10 to $1, some remarkably high ROI numbers. I cross-referenced that to the method that they use. Method one, with the variation, reports the lowest ROI. Personally, I think these ROIs are wildly overstated. These are public statements from disease management companies about how they calculate their ROIs, so I think that you could see the gap between rigorous method and what people feel comfortable reporting publicly. Probably the two biggest problems that I see coming up again and again are regression to the mean and selection of who winds up in disease management programs.

Most disease management companies have been offering services for less than 10 years, most of them for less than five, so it is a fairly new service. There is no one type that predominates. There are hospitals that offer them—HMOs, private and public companies. The staffing ratios are very dependent on the enrollees and on

the stratification of the enrollees. Again, I did not see any commonality. I think that this has a lot to do with the immaturity of the industry, because when you are a start-up company, when you have your first disease management enrollee, you have to have your first nurse at the same time. I think that we saw really low staffing ratios because so many companies were startups.

Information technology and medical management systems appear to be key. It is almost impossible to run these programs out of a shoebox full of index cards. You have to have some way in which to identify the member and get him or her on a track of education and regular intervention. You have to cue the nurse to make a callback or do another mailing or talk with the physician. And when incoming calls come from either the physician or from the member, there has to be some way to rapidly retrieve that information. You can't go out to a chart room or flip through a cardboard box to try to find that information. You have to get your hands on that information immediately. This is a chronic illness. It isn't a one-time event. You're talking about a progression. I also found that companies typically do not insure or reinsure their performance guarantee. So even though they would guarantee financial performance, they typically do not insure or reinsure that in any way.

Some important considerations—some of these programs graduate people, and some of these programs do not graduate people. Some of the programs will say that at a certain point, typically it is somewhere between six and 18 months into the program, the patient can manage himself. He is taking his medications regularly. He knows what he has to do to take care of his body. He knows what he has to do about his doctor's appointments. The program has grown his own knowledge and ability to care for himself. He graduates from the program. If he slips back or develops an exacerbation or has more problems, he can always be re-enrolled.

Other programs do not graduate you. They basically say that once a patient has a chronic illness, that patient is always chronic. They no longer can take care of themselves. They are going to need somebody running alongside them every step of the way. I think that people should be graduated to self-management. People should be able to manage their own medications, their own interventions. I actually think that its arrogance on the part of the medical community to think that people could not.

Information-technology (IT) capabilities drive the analysis. When you are looking at a disease management program, you really need to look at the company's ability to analyze its own information, both to find members and to design interventions. So IT capabilities are key to the analysis.

Interactive medical management systems that cue the nurse are very important to these programs. Several programs make a big deal about the Web services that they offer. There are two kinds of Web services. They can offer services for enrollees to go online, look up information, ask questions and get them answered,

etc. And then there are systems that are designed for the doctor. The doctor can go online, look up information about a patient, find out information about the disease or new medications, etc. It does not seem to offer any particular advantage, but it seems to be very tied to doctors and patients in whatever area of the country they are or whatever group they are working with prefer. So it could be a huge attractor for some and mean nothing to others.

HIPAA compliance is a new issue in disease management. HIPAA guards private health information. Disease management companies are now asking for and collecting that information, as well as collecting claims information that contains that information. There are a lot of questions about how the disease management vendors then work with doctors and the insurance company. We do not know the answers to these questions yet, but it is an issue for disease management companies that has lots of implications for enrollees and enrollment.

Professional liability insurance for providers is coming into question, too. When you are a disease management provider, are you delivering care? If so, do you need malpractice insurance? It turns out that a fair number of disease management companies do not cover their doctors and nurses that are employees of the company with malpractice insurance.

And, lastly, National Committee for Quality Assurance (NCQA) and URAC are new accreditations. NCQA is a big health plan accreditation organization. URAC is for utilization-management companies. It is becoming a mark of quality to have these certifications. Both NCQA and URAC have new certifications out for disease management that appear to be not as rigorous as their case-management and utilization-management programs, but that represent a great start as far as what programs should look at.

MR. DUNCAN: I am going to present results from programs that we have run in the past. I am in the process of analyzing this data. One of the things that distinguished our programs from others out there was that we ran randomized control groups; for example, the assignment to intervention was based on the last couple of digits of the Social Security number. There was no playing games with who got into the control group and who got into the intervention group.

To me, this was always the gold standard of outcomes. Using our results, we were able to show the client that the control group had significantly higher outcomes than the intervention group, post-intervention. There is a lesson in this, however, because the client was unhappy with these results.

The client's issues had nothing to do with measurement: we could show that we were controlling costs and that people were having better outcomes than the control group. But the health plan's trend was continuing to increase at 14 percent a year. Whatever we were doing on our intervention group was not noticeable in the outcomes that the financial management of the health plan was concerned

about. And that, frankly, became a very significant issue and led to the end of our program. So even with solid measurement and an agreed measurement methodology, the client will not necessarily accept the results.
Chart 14 shows a typical set of outcomes from a peer-reviewed, published study. People who enrolled in the program were compared to people who did not enroll in the program. People were offered a program depending on their diagnoses. Some enrolled. Some didn't enroll. Over a period of time, the costs of people who enrolled were compared with the people who did not enroll. Everybody's change in cost was negative. For example, the reduction in cost for people who have had an asthma diagnosis was -$480 over a year. For people who did not enroll in the program, it's -$75.

The hypothesis is, this disease management program has had significant savings results for the client. One of the things that we did with our randomized data was compare people enrolled in programs and their outcomes, with the outcomes of those who did not enroll in a program. What kind of outcomes do you see in people who do not enroll in programs? Chart 15 shows the result of selection bias. This is all measured in quarters. The point of intervention is the zero mark. And we went back four quarters, because we were identifying people who were at risk of having high costs in the future. Significantly higher cost outcomes, in this case, measured through a proxy of bed days per thousand per year were experienced in the group of people who did not enroll in the program.

Chart 16 further breaks down the data regarding people who don't enroll in the program, separating them from those people who refused enrollment in the program. And then people who you can't reach at all, who you don't have a good phone number for or who perhaps are already in the hospital when you try to reach them, actually have the absolute worst outcomes of any subpopulation. This seems to suggest a very significant selection bias present in enrolled versus non-enrolled study. So you should be cautious about any measurements that are based on that methodology.

I don't, unfortunately, have any data on this. It is too new. But the American Healthways methodology of taking a group and then projecting it forward, with adjustments, seems to be becoming the most popular method in the industry. Some people refer to it as the actuarial method, which is good, because it creates opportunities for us. This method has certain issues and certain problems. The biggest one is that if you are going to measure something a year ago, you need to project it forward with trend. We all agree on that. We need some sort of trend assumption. You can't use the trend that you observe in the population that you're managing because you've already intervened on that.

Where do I go for another trend to apply to this population? One of the suggestions made is to analyze the population that does not have the disease. So we'll use that as an index population, it's a population that has no disease. In one particular population, what trends do we see in the population that does not have one of the

marked diseases? Over a two-year period, the range of numbers was anywhere between –3 percent and 30 percent. With that kind of range, these companies are going to be able to prove whatever they want to prove.

MS. MARLA PANTANO: (Aetna.) One of the things that Aetna is considering is not only health-cost reduction but also some of the non-health reductions, such as in disability or returning to work. Has anyone looked into that?

MS. JOHNSON: I am aware that several companies are using workers' compensation and disability information, but I didn't come across anyone that was doing rigorous measurement of that, yet. Interestingly, some large industries (especially companies that experience a lot of things like low-back injuries that are amenable to disease management) seem more likely to want to use those measures. Most of the programs for disease management are for things like asthma, congestive heart failure—chronic medical conditions that are unlikely to have been workplace-induced. That overlap isn't great.

MR. WILSON: If you can measure it, we can use it. Measuring productivity is difficult, especially if you're using absenteeism. A lot of absenteeism has nothing to do with sickness. So it is an issue of measurement. And I like the HRAs. But of course, you cannot use them much, because they are expensive. For a baseline measure, it is good. We do need more information like that.

MR. DUNCAN: If companies cannot justify their programs prima facie on the results of their programs, then they shouldn't have those programs. If you have to grasp around for justification with second-order effects, like time off and reduction in disability costs, then you are admitting that the programs cannot be justified based on medical cost savings alone.

MR. WILSON: Can I disagree with that just for a moment?

MR. DUNCAN: Absolutely.

MS. JOHNSON: And I would disagree, too.

MR. WILSON: Who gets the value from improved health? Is it the health insurance company or is it the employer group or whoever? Whoever gets the value should be the one that's paying for it, in my opinion.

MR. MICHAEL DUNN: (AON Consulting.) Obviously, the clients want to feel that they're getting some value from these programs and that their investment is worthwhile. There are a couple of issues that I have encountered in a selection process with disease management vendors. One is in the identification process. Many vendors claim that they can integrate the health risk assessment data, the medical claims data, the prescription drug data. Are there companies out there that really have proven their success with effectively identifying and linking all of those data sources, especially when they come from multiple vendors? And on the

enrollment side, virtually every vendor that I've encountered in a "finalist's presentation" has stressed the need for some kind of financial incentives to drive participation, both in the health risk assessment phase as well as participation in the intervention program.

MS. JOHNSON: The financial incentive to fill out a health risk assessment is one that comes up continuously with all kinds of debate. I didn't even collect the list of the various things that people were offered. But if you have an enrolled population—and you are just working off of the enrolled population rather than re-enroll people in disease management—they will mine the existing data and then go to the enrollment file. So they will not have to re-collect information from people from that standpoint. Does that make sense?

MR. WILSON: On HRAs, I am actually presenting with a group at a conference in Boston on the linkage of claims, pharmacy and HRA data. The issue is congestive heart failure. We are looking at the added effect of having dementia, as ascertained by claims data and health risk assessment data. It is not finalized, but it appears that for undiagnosed dementia, in conjunction with congestive heart failure, there is a much higher cost. If a physician would diagnose dementia and treat it, it would have a big impact. Anyway, we are doing that kind of work, linking all of these different data sets, using this epidemiology data system.

MS. JOHNSON: I did talk to a gentleman who had done some work with organizing banks so that you can put your money in the ATM anywhere and get local currency. He decided to try to integrate health care. He said that trying to work with the various health-care systems was much like trying to work with the patrons at the bar in the movie "Star Wars." He said that it is not even the same species. It is a tremendous problem, trying to integrate that information. And the worst ones to integrate are your basic medical claims, pharmacy and behavioral health. It looks as though it is the biggest bang for the buck, and we cannot even do something as simple as integrate those three.

MR. DUNCAN: I would voice an opinion on paying people to do HRAs. One of my clients had a company that paid employees $500 to go do an HRA. That's a very attractive amount of money, and they got a good response to their HRA, but for the kinds of savings PMPM that you can get out of these programs, there just isn't the economics there to support paying people that kind of money to do an HRA.

MR. WILSON: I was involved in a program for which they wanted to give people an incentive to do HRAs and then promote good behavior that resulted from that. The year that they did the incentive, the smoking rate dropped like 50 percent, but that is because people lied. The number of people doing physical fitness increased by 35 percent, because participants got paid to say that. So we've got to worry about those things.

FROM THE FLOOR: Can you get better ROIs, depending on the structure that the management company has developed to work with providers?

MR. WILSON: First of all, there are not any really good ROI studies, which is what your question is predicated on, but it's an excellent question. Once you do it in a credible way, and I think that it has to be done in multiple time segments, in multiple markets.

MS. JOHNSON: Is it dependent on the provider, or not? There is a real debate about whether or not you really need to get to the doctor in order to be able to get to the patient, or whether getting to the patient is enough. Is it the responsibility of the patient to manage his own illness, take his own medications, follow up on his own things?

Then the second thing that I want to point out is that a recent study from the Pacific Business Group on Health analyzed various medical management activities and the ability of nurses that work in those programs to influence primary care physicians. And the basic conclusion is that physicians ignore those telephone calls from nurses. It is just about impossible to integrate that into the daily work of a clinic. The doctor is seeing patients every 15 minutes. And when he gets a phone call from a nurse about some patient who isn't even in the clinic, he has to go retrieve the chart, remember what's going on, integrate new information into this, etc. Then he has to make an outgoing call to the patient. And there is no payment to them for that, because the patient is not in the clinic. The impact of nurses in disease management and medical management on practicing physicians is very tiny.

FROM THE FLOOR: It seems to me that one of the debates that exists in the DM industry is whether or not lasting behavior change actually occurs with these individuals. In fact, you can run the gamut from an information-only program to a handholding program. There are precious few programs that actually get into cognitive behavior change activities, with individuals intervening to do that. Did you look at that particular information? You talked about moving people to graduation. That would suggest, if you move them to graduation, that they're going to be motivated to do that, that they're going to have a cognitive behavior change. If that does not happen, what have you done? You have a program that continues to be in place with the same people and just keeps working like a mill.

MS. JOHNSON: That is a great point. It really is predicated on the person being able to make a permanent change. There are some changes that work and some changes that don't work. We know, for example, that problems with smoking and obesity are very difficult to change. Other things, like getting people to adjust their medications, seem to be much easier to do. For some of the things, the impact on the person is immediate. For example, if you have a two-pound weight gain, you are on your way to the emergency room within the next couple of days if you have congestive heart failure. And it does not take too many emergency room visits for

the person to figure that out and be very motivated to try to prevent things from getting out of control. For other patients, such as diabetics, who have to test their blood sugar every day and make adjustments in order to prevent blindness in 20 years, such things are very difficult for people to do. In the research study that I did, I did not look at those factors. Most disease management programs incorporate some sort of a change model that is the foundation of the nurse's work with the member. But are they successful? The jury is out.

MR. WILSON: I do not think that there have been any good studies of that particular issue, except in smoking-cessation programs. If you stop smoking, do you go into remission or not? What is the success of that? And all of these things go back to the issue of what would have happened to your smoking-remission rate had you not had this program? I smoked 20 years ago, and I never had a program. I quit, and I am in remission. Who should get paid for that? This is the big question that people are asking about all of these programs. I know that the federal government is extremely interested in disease management, because it really is the only thing left for managed care. It's the best new idea. Nobody has any other idea. The idea is that we need to improve people's health in some capacity, compared to what it would have been otherwise. We cannot do all of these strict precertifications anymore. Getting a handle on that is the future of this industry. It is really *the* critical issue. And behavioral change—permanent, temporary, how do you know? That's the question that we're going to have to start answering.

FROM THE FLOOR: It seems to me that the issue with DM is, is it going to be a lasting thing in terms of behavior change? I would suggest that that is where the focus has to occur with the programs that exist right now. I don't care what their results show and what method they use today. Has it created that change for that individual on an ongoing basis? That point would be relative to anybody evaluating that kind of program.

MR. WILSON: What I would do is enroll people, and kick 25 percent of them out at six months, 25 percent out at 12 months, another 25 percent out at 18 months, and the rest at 24 months, and then see what happens to them.

MR. DUNCAN: One of the theories that we had in our company was the whole direct-marketing approach. We believe that people need frequent but small reminders. And this did not go over well with health plans that do not want you to invade people's privacy and pester people to change their behavior. On the other hand, if you think about what people are bombarded with on a daily basis in terms of messages to go and buy French fries, it might be necessary. The health care industry is not willing to be as aggressive about trying to change people's behavior as those people who have a more financially vested interest in it.

FROM THE FLOOR: It does not sound as though there is any real data out there to substantiate disease management. What are three good questions that you would ask vendors to try to determine if they are ethical companies or not?

MS. JOHNSON: I would ask them if they published any research about their results and ask if they could share all of that with you. I would ask them what they're doing about enrolling and stratifying people and have them talk about their concepts. Then I would ask them about graduating people from the program and what their beliefs are about self-management. Those three questions might help you find out if they have the best interests of the patient at heart. And are they aimed at producing a financial result that would be meaningful to you?

Chart 1

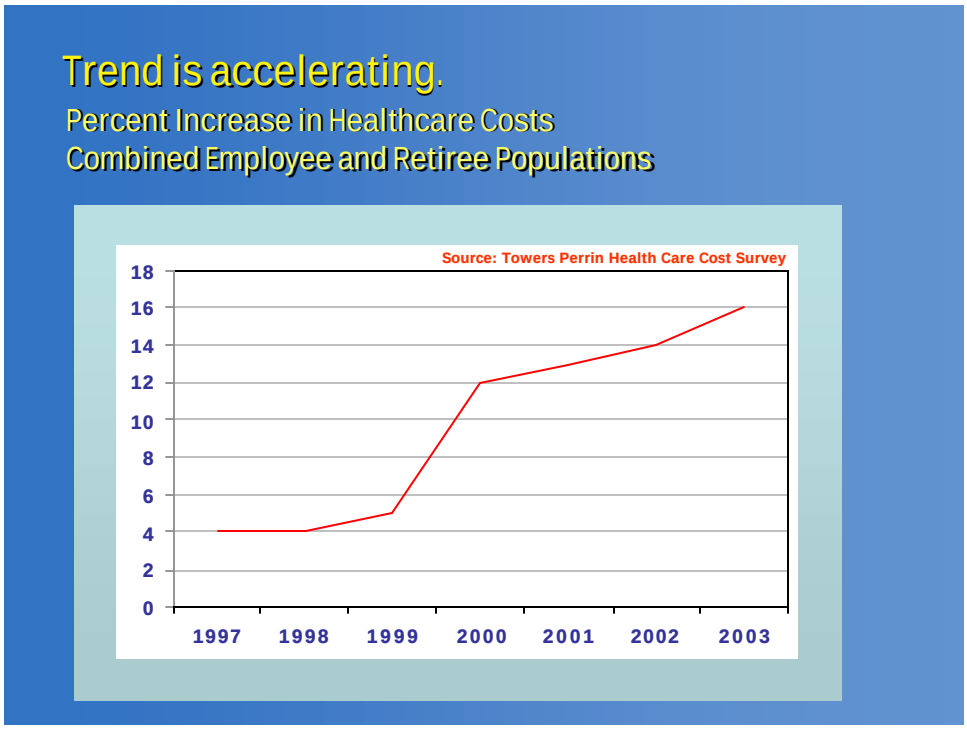


Chart 2

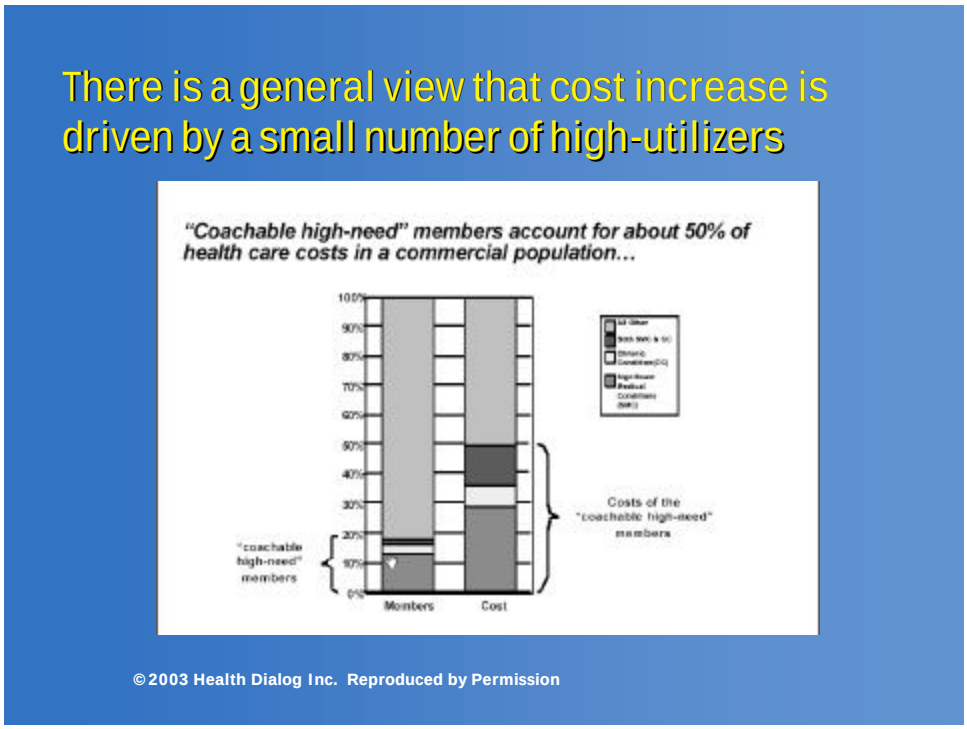


Chart 3

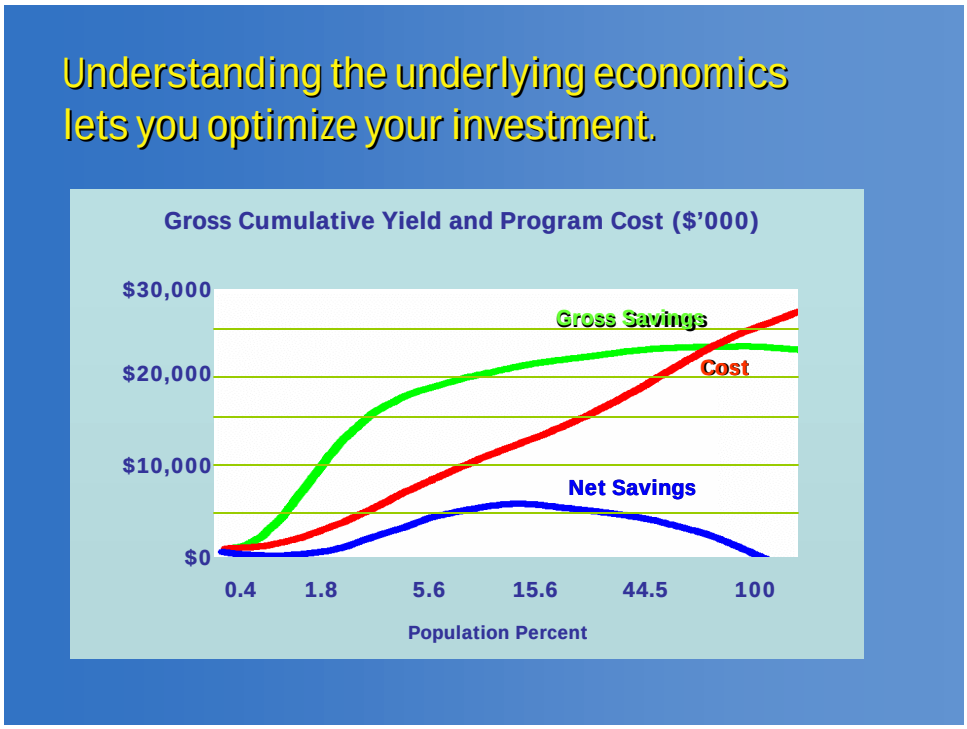


Chart 4

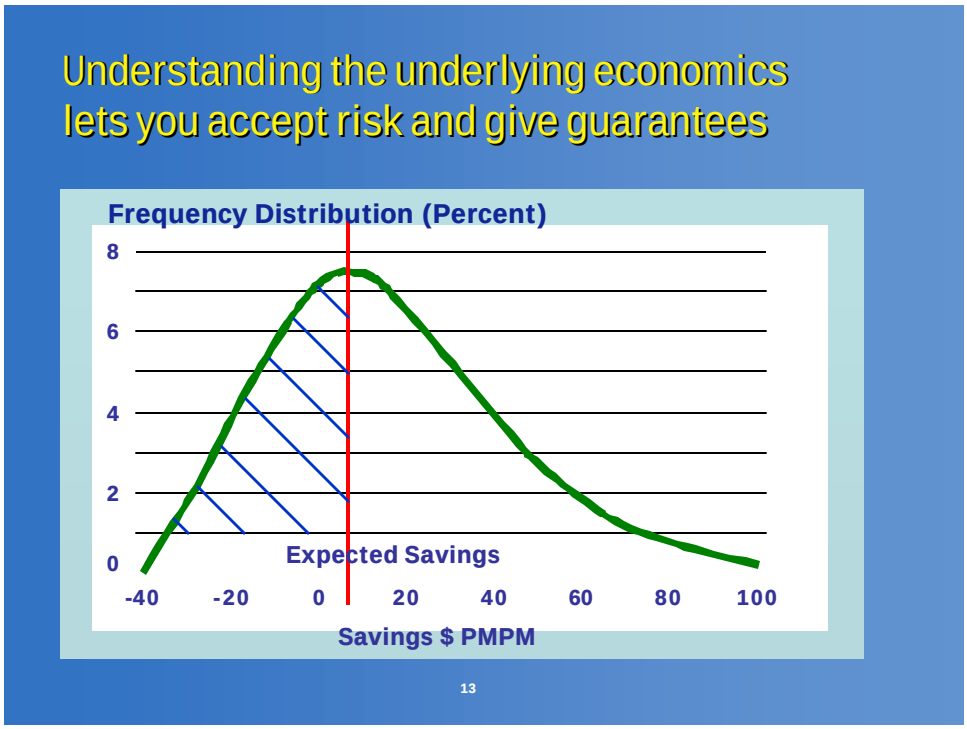


Chart 5

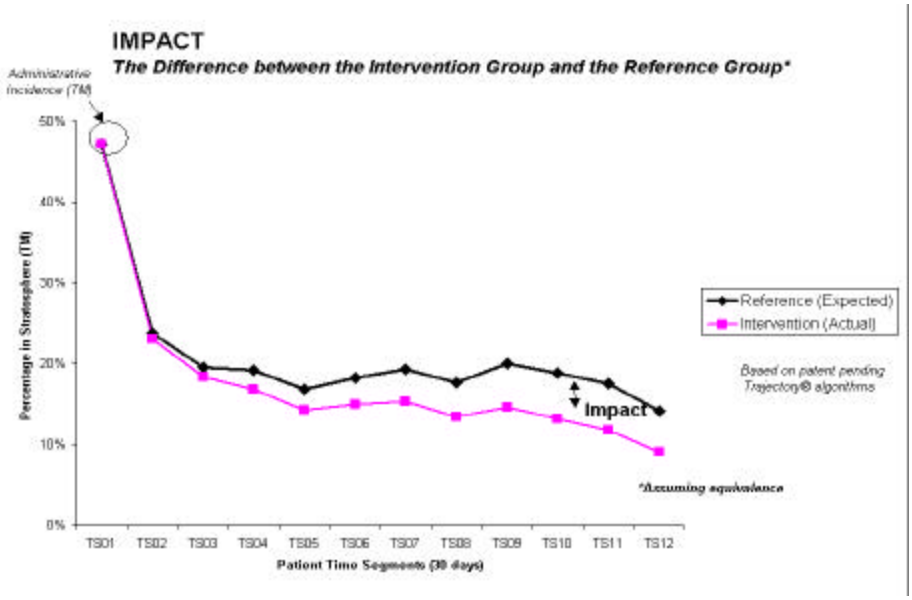
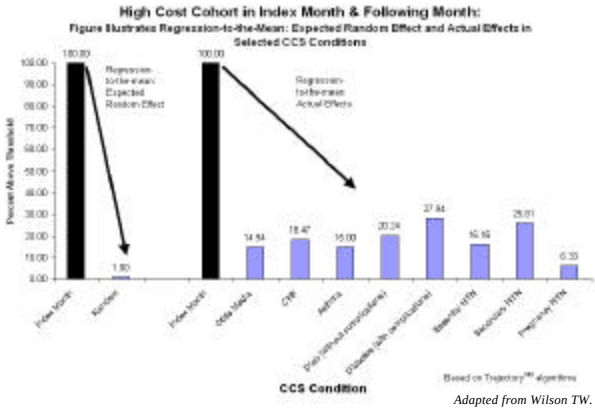


Chart 6

"How I Learned to Stop Worrying and Love Regression to the Mean"

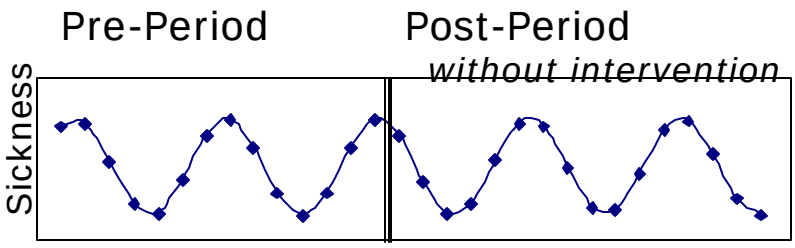


Adapted from Wilson TW. "How I Learned to Stop Worrying and Love Regression-to-the-mean." Presented at American Association of Health Plans--Building Bridges Conference, April 2002.

Chart 7

Pre-Post Design

Equivalence Assumption

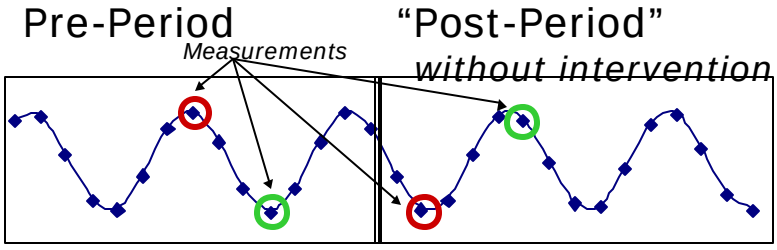


Thus, in a properly conducted pre-post study, any change detected in metrics in the post-intervention period could, arguably, be attributed to the DM intervention.

Chart 8

Pre-Post Design:

Past is NOT Prologue: *A Situation where equivalence is not achieved (if you're Red or Green)*



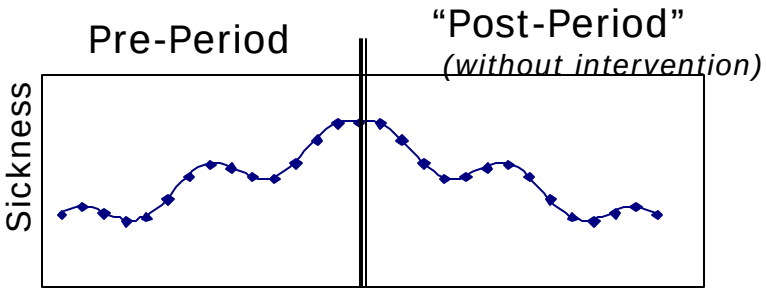
Spurious Progression: ***Measured at low end of cycle in pre period and high end in post period***

Spurious Regression: ***Measured at high end of cycle in pre period and low end in post period.***

Chart 9

Pre-Post Design:

Past is NOT Prologue: *One Situation where equivalence is not achieved*



Not a good situation to conduct a pre-post design unless you are aware of this trend and take it into account in your results.

Chart 10

"Past is Prologue"

Two Situations where equivalence is achieved (as long as you are aware of it)

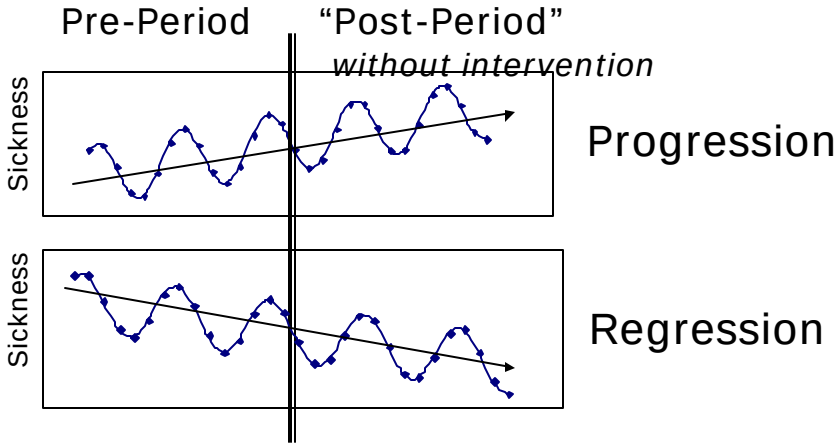
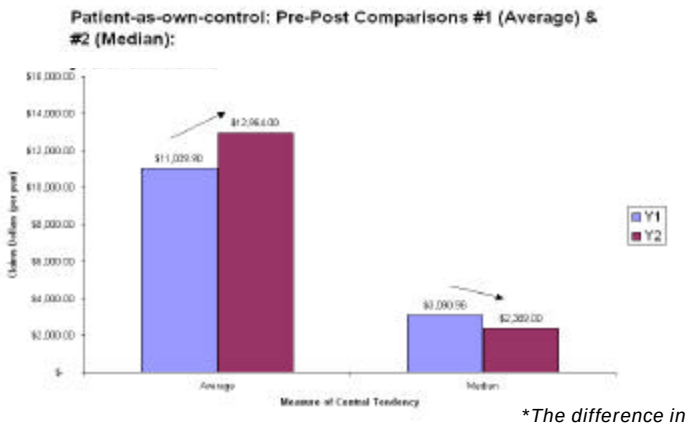


Chart 11

Patients as Their Own Control:

*Averages vs. Medians**



**The difference in outcomes is due to skewness of distribution of cost variable*

Chart 12

Time Series:

Example where equivalence assumption is problematic

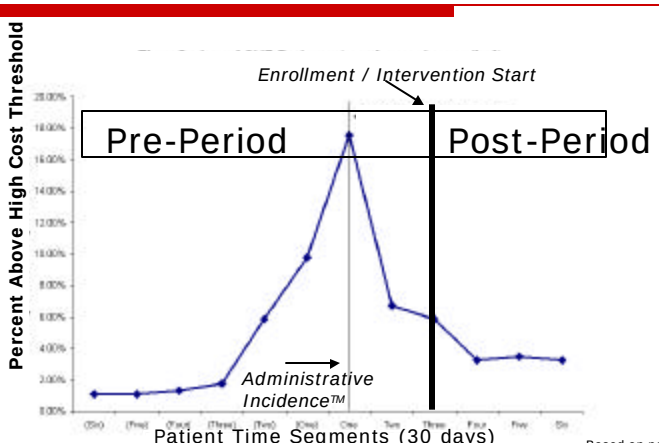


Chart 13

VII) Follow-Up Design: *General Characteristics*

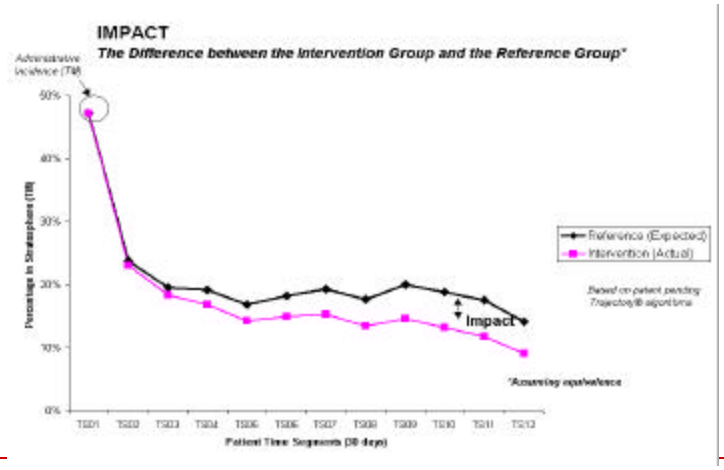


Chart 14

Example 1.

- **Enrolled vs. non-enrolled**

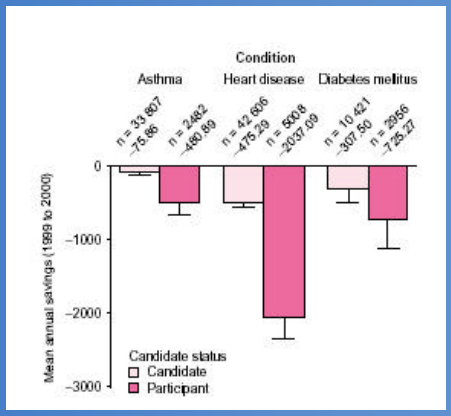


Chart 15

Example 2.

- **Enrolled, non-enrolled and Randomized Control**

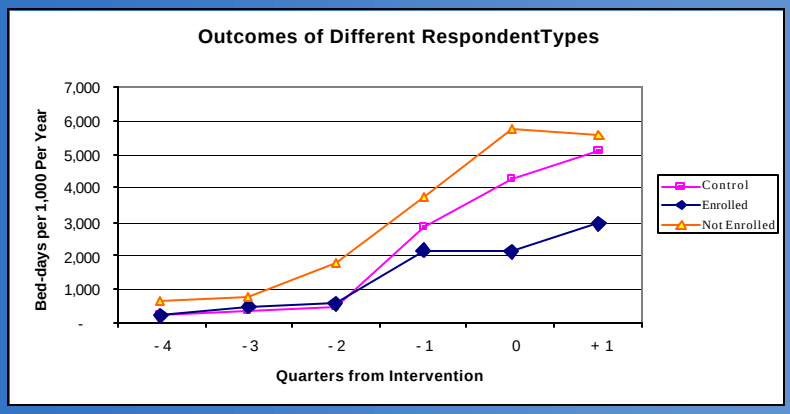


Chart 16

Example 3.

- **Enrolled, non-enrolled and Randomized Control**