RECORD, Volume 31, No. 2*

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

Session 31 Seminar
An Introduction to Care and Disease Management Interventions Part 1

Track: Health

Moderator: IAN G. DUNCAN

Panelists: IAN G. DUNCAN
HENRY DOVE†

Summary: In 2003 the Health Section of the Society of Actuaries, recognizing the increased need for rigorous calculation of the financial outcomes from care and disease management programs, sponsored an extensive research project into the actuarial issues of these programs and their financial measurement. The study encompassed both theoretical and practical aspects, including analysis of outcomes from an extensive disease management program that has been in place for a number of years at Highmark, Inc. Part 1 of the seminar examines some of the key issues and findings from the research, including disease management program types, program experience and program results; literature review of care and disease management program outcomes; understanding the economics of care management programs; and outcomes and outcomes measurement methodologies. A related paper from this study, "A Comparative Analysis of Chronic and Non-Chronic Insured Commercial Member Cost Trends," is scheduled to be delivered as part of the Financing Chronic Care Seminar.

MR. IAN G. DUNCAN: My colleague Henry Dove and myself are going to be presenting the results of a two-year study that we've been engaged in and that the Society of Actuaries' Health Section has sponsored looking at different aspects of a growing area of interest for actuaries as well as medical management companies

*Copyright © 2005, Society of Actuaries

†Dr. Henry Dove, not a member of the sponsoring organizations, is lecturer at Yale University Health Management Program in Hamden, CT.

and health plans, that of disease management and other care management interventions.

I'm pleased to be an actuary. I'm interested in the future of the profession and its opportunities. I have two offices, one in New York and one in Hartford, Conn. Henry will introduce himself in a moment and give you some of his background. He is not an actuary. We're going to talk some about the research that we did. I'm going to talk a little about the process because I found this quite interesting. I had not done research before, either in an academic situation or for the Society of Actuaries, so I learned a lot about the process along the way. Research is not something that within the Health Section of the Society that actuaries get into much, so there are some learnings there to pass on.

We're going to talk about the environment and the background on care management and about some of the current research going on in the industry. A total of eight papers have been written and submitted to the Health Section of the Society. I think five of them are to be found on the Health Section Web site at the moment. The rest are going through peer review and will appear later.

Let me give some background to our research. Traditionally, actuaries have been involved in financial rather than clinical topics. Managed care brings two streams together: managing clinical interventions for a financial outcome. Managed care is obviously there for a clinical outcome, but managed care is there in part to produce a favorable financial outcome as well. Actuaries have begun to get more involved in the care management and outcomes side of the business as a result of this. This creates a need for actuaries to learn a whole new vocabulary. When I came into this business a few years ago, I had absolutely no idea what people were talking about and had to learn it all from the ground up. I had no training at all. It also creates opportunities, though, for actuaries to carve out a niche and advance in a different area. I think this is important for the profession as we move the image and move opportunities for actuaries forward.

Back in 2003, almost exactly two years ago, the Health Section of the Society called for proposals of projects that people were interested in doing. I had been running a small disease management support company. It was a start-up in 1996, sold in 2000 and then eventually shut down in 2002 by its new owners. As a result of that experience, I got very interested in predictive modeling and in disease management, particularly the measurement issues around disease management. I had some interest in exploring this further. Henry and I submitted a proposal to the Health Section, which was accepted with some changes.

It was interesting what the Health Section wanted to change about the proposal. We had originally seen this as being a paper on the actuarial issues to do with measurement, perhaps a paper on actual applications to data and focusing on outcomes. The study grew from there to being a total of eight papers, in part because the Health Section oversight group that was looking at this said, "We want

you to do an introductory paper, a first paper, that simply talks to people who know nothing about the subject of care management interventions." That was not something I had thought about, so that was a good idea. The group also insisted that we do a literature search. This is where my complete absence of experience in academic pursuits showed up, because it never occurred to me to go out and look at everything that had been published before in this area. We reviewed about 2,500 papers in total that had been published in the peer-reviewed literature on different care management interventions. We whittled the 2,500 down to about 100 that were of interest to us because they had financial outcomes, and then summarized them in the paper. A couple of other topics cropped up along the way. One is the whole area of the economics of medical management, which I think is something that is given very little time and attention and needs a lot more. We've also done a paper on trend, which we'll talk about briefly today and which will be presented in more detail at the chronic care financing seminar tomorrow.

So a total of eight papers have been published or are in the final stages of preparation. They've gone through a *rigorous* peer-review process. A couple of years ago Henry and I published a paper in *The American Journal of Managed Care*, which is a peer-review journal in the health services research literature. The peer-review process was reasonably onerous. People sent back comments and we had to go around a couple of iterations of rewrites, but eventually it was published. It took six months and away we went. These papers have been reviewed by the Project Oversight Group over and over and over again, much more rigorously than anything I had ever anticipated. The Project Oversight Group has put almost as much work into this project as Henry and myself. There has been a tremendous commitment of time, energy and effort on the part of volunteers in this organization. When you're a health actuary and you're practicing in a health plan, you don't think about what's going on back at the Society on the research side. We can only say that we owe these folks a lot.

The eight papers that we'll talk about in the next couple of hours are: an introductory paper on programs and interventions, a paper on actuarial issues in care management, a paper on the review of the literature that I talked about, a paper on the economics, a paper on the different methodologies for measuring outcomes, a paper on actuarial methodology for assessing disease management outcomes, a brief oversight of the paper on trend (that's the subject of another presentation tomorrow) and then some practical applications with some real data, which is what you're all waiting for, but you'll have to wait until the end.

A large number of people need to be thanked and credited for their contributions to this project. It was a two-year project, and a lot of people were involved. My co-author is Henry Dove. My other co-authors were Rob Bachler and Iver Juster. We owe a tremendous debt of gratitude to Highmark Incorporated and Bill Cashion, the chief actuary, who made available several years' worth of Highmark data and the resources to do the analysis of it. I'd like to thank my colleague Rebecca Owen, who started off with Highmark and now works for me in Hartford, for working on

this writing and analysis. The Project Oversight Group, which consists of Bryan Miller as chairperson, Marjorie Rosenberg, John Cookson, John Stark and Stacy Lampkin (all very well-known names), contributed an immense amount of time and effort to this enterprise. Ronora Stryker, the Society's research representative to the project, provided much support and encouragement. Last but not least, I'd like to thank the SOA Health Section and the Committee on Knowledge Extension Research for their valuable support, which actually came with a little bit of money attached. There was enough money attached, I think, to make it at least worthwhile thinking about doing beginning the project, and to fund cheap resources like students we hired from Yale. There are ways to do research, and I think the seed money from the Society is quite valuable. We need to think about other projects where some seed money could produce useful returns to the Health Section.

We're going to do a session tomorrow on health actuarial research. If you look at the *North American Actuarial Journal* (*NAAJ*), out of about 250 articles that have been published since the *NAAJ* started maybe nine or 10 years ago, something like 15 percent of them have been about a health topic or have "health" somewhere in the title. (I drew the net widely and included things like genetic testing, for example.) Yet if you look at the people who put "health" or "the Health Section" as their primary interest in the Society, we're about one-quarter or one-third of the members. So we are woefully underrepresented in research in the Society publication. If the profession is going to be taken seriously within the health sector, going forward we need to change that. We need to do more research, and we need to publish more if we want to be taken seriously and improve that image. We need to get out there, talking more and publishing more, because that's what all other disciplines in the health professions are doing.

Let me give a little background on care management. There's clearly a realization that a small percentage of health plan members consume a large percentage of resources. There's a growing awareness that the health plan member has some role in this, and so we think that we should be able to encourage the members to take more responsibility for their own care. One of the results of this is an interest in disease management, which is a set of interventions that recognize the role of the patient in his or her own care.

All these interventions that you practice upon the patient raise all sorts of interesting questions for actuaries. Here's the first one: Medical management departments are expensive resources. Actuarial departments are quite well-paid, but medical management departments probably outrank us on the pay scale. They're expensive resources. They tend to be under different management structures than actuaries, and it's very hard to measure their productivity and performance. They also seem to think frequently that they should get some kind of waiver when it comes to showing financial results. After all, they demonstrably "do good" for people, so why should they have to show a financial gain? As the number and cost of intervention programs have risen, managements have begun to turn to

their traditional financial advisors, the actuaries, for advice. But clinical metrics and evaluations are not part of the traditional actuarial syllabus. So what to do? Well, that's why we were interested in the subject, and we were awarded a research grant to do this, as I said a moment ago.

With the change to managed care and care management, there has been a change in focus. Traditionally, actuaries are focused on health *services*. But we think the focus is shifting to the member and the member's condition. For example, what's a reasonable cost for a member with a particular condition? How manageable is that member's cost? What is the trend in cost for, say, diabetes, and how does this contribute to overall health plan trend? These are the kinds of questions that are beginning to be asked. If you look at the traditional way, the traditional axis on which medical services were evaluated, we looked at different service types, such as hospital inpatient, emergency room (ER) visits and so on. But the axis has shifted, and in the future we're going to be looking at people first. That will be the primary axis for analysis. Then within that, we'll look at the services that people consume. One way to look at people is a chronic and non-chronic split, so a lot of the analysis that we do looks at people on a chronic and non-chronic basis.

There's a little bit of other research and progress in this area. The American Academy of Actuaries is working on a practice guideline for actuaries practicing financial outcomes measurement. The Disease Management Association of America (DMAA) has recently published two books. One of them is called *Dictionary of Disease Management Terminology*, and the other is *Disease Management Program Evaluation Guide*, which takes you through a question-and-answer format in how to approach some of these issues.

The American Academy of Actuaries came out recently with an issue brief on disease management that you should read. It's called "Disease Management Programs: What's the Cost?" It's about three or four pages long, so it's quite brief. It is the precursor to what I talked about a moment ago, which is the practice note that the work group is working on currently to provide guidance for actuaries who are working in this area.

Henry is going to start going through some of the papers now.

DR. HENRY DOVE: Let me first say a little bit about myself. I came into the health services research field following my service in the Army in the Viet Nam era. I worked for the Texas Hospital Association on a research project in the early days following the implementation of Medicare legislation. After that, I became a graduate student at Yale University. This was at an interesting time because it was during the development of the diagnosis-related groups, which was done by Professors Fetter and Thompson, who were my dissertation advisors. When I was a graduate student, my discipline was operations research, but I did a lot of work at the school of medicine, because I was interested in clinical decision-making and clinical epidemiology. That has influenced a lot of my thinking, which I think you'll

see in some of the papers that we've written here.

What I want to talk about in Paper 1 is the introduction to care management interventions. I know that this will be familiar to many of you, and I'll go over this part reasonably quickly. Ian and I came up with seven categories of care management: preauthorization reviews, concurrent review, case management, demand management, disease management, specialty care management and population health management. But this list is changing. I'm sure that if we gave this session five years from now, two or three new terms would show up. When I go back to 1968, right after the period of Medicare implementation, costs were going up 20 percent per year. The federal government was very nervous because President Johnson was instrumental in passing Medicare and Medicaid, but there was always this concern about there being no end in sight as to how high these costs were going to go. So what can we do about controlling these health-care costs of Medicare? That was the beginning of the managed care in a way, as we were desperately looking for ways to control the Medicare costs.

There were three factors. One was simply that the number of people 65 and older was increasing. There was also a big problem of increased utilization and a problem of increased prices, because every year the hospitals would raise their prices. This is what led to the research that began back in 1965, which ultimately lead to the development of diagnosis-related groups, which was implemented in 1983. As part of that, the hospitals were responsible for doing their own utilization review, which was a new term back in 1968. So what occurred back in 1968 was the introduction of various ideas to control the utilization of hospital resources back in those days. As managed care took over, they came up with their own ideas to control the price and utilization of medical resources.

The first technique that was used was preauthorization. Before you could have a service performed, you had to get permission from someone. Now I'm not thinking in terms of Medicare, but I'm thinking in terms of the private managed care organization. As you can imagine, this led to a lot of battles between the managed care organization and the patient and the physician. These had to be worked out. But then there was the question: What was the impact of the preauthorization reviews?

Another part of this was the admission review and concurrent review. The idea of this is monitoring a health care plan member's care while he or she is receiving care in an acute hospital or in a nursing home. This probably caused more controversy then any of the other interventions here. As Ian said, this was incredibly expensive. You had nurses that either worked in the hospital or went to different hospitals. This began the 800-number call-in; the hospitals had to call in this 800 number to get permission. Once the patient was admitted, there was the continued-stay review. The physician or the utilization review nurse said, "Well, we approve this admission and we want to review the care again one week later to see the progress of the patient and the discharge plans and so forth." This caused an incredible

amount of friction.

The residents and the interns became very used to some of these practices, and of course, the greatest resistance occurred from the older physicians, who were not used to answering to anyone with respect to their treatment plans or medical resources. But one of the things that happened was that the younger physicians learned what criteria would enable them to have the patients be approved for their admission and for their continued-stay review. There was this sort of dance that went on, which eventually reached a steady state, more or less. There was an enormous amount of resources that were devoted to preauthorization reviews and concurrent reviews, but there was not very much evidence whether or not they had an impact, either in terms of cost or in terms of quality improvement (clinical outcomes, basically).

A third technique is case management. This is where a health-care professional coordinates the care of a patient with a serious disease or illness, such as stroke, multiple sclerosis or AIDS. This has been pursued not only by managed care organizations, but also by specialty organizations.

The fourth technique that we studied was demand management. This was the beginning of disease management. Demand management was a more passive form of intervention. In the typical case, a patient could call in a nurse to get advice, with the hope that the nurse would keep the patient from unnecessary utilization of the emergency room or that the nurse could provide more information to the patient so that a patient didn't have to go to the physician for a visit. Another part of this is shared decision-making. This is an attempt to have more discussion between a patient and the provider to talk about different options and to have a discussion about critical decisions that have to be made before they are made. Shared decision-making takes into account more of the patient's preferences rather than just having the physician say, "I think you need surgery," or "I think you need to have some complicated diagnostic intervention."

There are three more recent managed care interventions. Disease management is involved with the management of chronic conditions, usually with a combination of pharmaceutical therapy and lifestyle changes. Another technique is specialty care management. This also involves a care manager that has particular expertise, such as in mental health, organ transplantation, oncology or end-stage renal disease. What is different about specialty care management is that frequently a managed care organization will carve out these patients to a separate firm that takes the financial risk for the management of these special clinical problems. The last technique is population health management. This is a softer form of intervention, which involves the use of predictive modeling. That is the most recent development on the horizon with respect to managed care interventions.

We were the ones that came up with the list of those different categories. Sometimes it became difficult when we looked at the literature. As we found an

article, it wasn't clear which of these buckets to put those into, so we certainly can't claim that our list is the best list or the only list. But we had to start somewhere.

The second paper has to do with actuarial issues in care management. The idea behind this article was that we had a discussion of measurement principles, different study designs and issues about risk factors. I had a long-standing interest in risk factors because of my interest in severity adjustment as part of my dissertation. I think that actuaries have a slightly different, but similar, notion about risk. For me, the notion of risk, coming from a more clinical background, had more to do with risk of dying or risk of attaining a certain clinical outcome, whereas actuaries think of risk as chances, or probability, that a patient is going to incur very high medical expenses. The basic idea is the same; it's to look at certain independent variables that may affect the probability that a patient is going to incur very high expenses or die or achieve a certain clinical outcome. Another part of Paper 2 is that it addresses some of the issues of particular importance, including the notion of regression to the mean, risk adjustment, the need for control and reconciliation of data and operational issues.

One of the things that the Society of Actuaries wanted to know was, after all is said and done with these various interventions of managed care organizations, do they save money or not? That's sort of a simple question, and of course, the answer to that is not simple. There are many different ways that you can address the question. We say that the financial jury is still out.

In the basic approach to managed care interventions, there were always the issues of access to care, quality of care and cost of care. When people were interested in, say, admission review, they weren't thinking initially so much about whether to save money. They were more interested in questions like: Is this going to harm the patient to whom you denied admission to a hospital? The actuaries, I think, are more interested in financial issues rather than clinical issues. This is something that came up as we did our work here.

You will find that there are a lot of unrealistic claims being made about the savings of the different interventions. A lot of this has to do with the methodology that is used. It varies all over the place, which is one of the reasons why it's difficult to evaluate these interventions. There is no standard way for assessing the financial outcomes. That is something that Ian and I have talked about. Ian will talk about that in particular in Paper 6, where he gives the actuarial approach to evaluating the value of interventions.

Another aspect of these interventions is that there is poor reconciliation. This is one of the reasons that it's difficult to evaluate whether interventions save money or whether they do not save money. There's a lack of understanding of the key drivers of financial outcomes.

Then there's the question, "So how come, if you say that you saved me all this

money, my trend is continuing to increase?" Ian comes from a disease management and predictive modeling and population management point of view from his work with a disease management firm, a population management firm. They were asking this same question back in the early days when the focus was on preauthorization review and admission review. It's just that they never had time to address this question. They simply said, "We're spending millions of dollars on hiring these positions and these nurses." They would do certain internal studies that would show their admission rate had come down, but nonetheless their costs were still going up. What is going on here? It's sort of like a moving target that we had to evaluate.

I want to talk briefly about Paper 3, which is a review of published peer-reviewed articles. We're giving a session tomorrow (Session 55) about how to assess a published study. We're trying to give the audience a motivation for trying to publish more articles. We used two basic tools to access the literature. The first is a wonderful tool called PubMed, which is available at the Web site www.ncbi.nlm.nih.gov/entrez . But the easiest way to find that is just to get into Google and type "PubMed" and it will take you to this site, which is the effort of the National Library of Medicine of the National Institutes of Health. This is a wonderful resource for doing literature searches. This is an extremely powerful tool. I've only been a user of this for a couple of years; before this I used Medline. PubMed is much better than Medline, which was also free and which also came out from the National Library of Medicine.

Another tool that Ian told me about was an effort by the folks at Google. This is called http://scholar.google.com. I'm not as familiar with that. I was impressed with what I found, but I think that more resources have gone into PubMed. I do not underestimate the folks at Google, so that's something you would want to watch.

I also want to talk about DMAA's outcomes database, DMLitFinder. This has a slightly different focus than what we were interested in. DMAA has, as you could imagine, more of an interest in clinical outcomes rather than financial outcomes. They have more of an interest in specific diseases that are commonly the target of disease management companies, such as congestive heart failure, asthma, ischemic heart disease and so forth. DMAA's database, for which you have to pay to access, is geared more for the disease management firm rather than what we were trying to do in our research.

How did we go about identifying articles that we reviewed in order to estimate the return on investment of their risk, managed care interventions? First we developed very liberal criteria to identify more than 2,500 different articles that involved care interventions, and then we eliminated articles without an abstract or that were published before January 1, 2000. It is astounding how many articles and how many different medical journals exist today. I'm convinced that anybody can publish anything if you just have enough energy to submit it to a variety of journals, because there are something like 3,500 different medical journals now.

They have mushroomed like crazy. To keep up on the medical literature these days is a real challenge, to say the least.

The second stage was that we reviewed each article's abstract and then eliminated those without financial or utilization outcomes. Our interest was not whether a particular intervention improved the clinical outcome. We took a financial focus to our study, based upon what we were contracted to do with the Society of Actuaries. Amazingly, we went from about 2,500 articles to a total of 106 articles.

In the third stage, we obtained the full-text version of review articles or metanalyses. Finally, we only used roughly 100 articles in which financial or utilization outcomes were an important component. This was a lot more work than I thought it would be when we started this project. As Ian said, we did use three Yale University students just to help us track down these articles, copy the articles and organize these articles.

One of the other trends that you see—you'll notice this when you use PubMed—is that increasingly there are full-text versions available for an increasing number of journals. The better the journal, the more likely it is that you will be able to get the full text of an article in that journal. I predict that five years from now, if you want to do a literature search, the challenge that you will have is that there will be more journals and the journals are getting longer, but you will be able to download the full-text version of all of those articles. It took us a long time to track down these articles, and the Yale University Medical School Library was helpful in getting us some of these journals. You can imagine how expensive it is for a medical school to subscribe. They can't subscribe to all these journals, and so we had to use special tools in order to get the full copies of these articles.

The "MESH" term is an abbreviation for "medical subject headings." This is the primary way that we identified articles. The terms that we put in were: care management, disease management, utilization review, economic evaluation, utilization management, case management, predictive modeling and cost control. I wish that I could say that at the outset of this project we could identify exactly the method that we would use to identify the article. But there was a lot of back and forth because in order to identify a certain article, the term that you used in order to identify the article was a critical factor. We knew that there were certain key, or seminal, articles, and we looked at the MESH terms associated with those important articles. Then we used the MESH terms that were used for those articles in order to find other articles.

We used the same search eight different times, and we came up with about 3,000 different articles. Since it was possible for one article to appear in more than one of the eight searches, it was around 2,500 articles that we went through.

I mentioned DMAA's LitFinder. This is much smaller; it has fewer than 1,000 different articles. They go into a review of these articles much more carefully, but it

wasn't especially helpful to our research. I do think we found a couple of articles that we might not have found otherwise.

Stage 3 was the review of the full-text articles. I want to talk a little about the review articles. Sometimes you would find an article, say a "review" of different interventions such as asthma. The author did not do any original research; his job was simply to put together all that was known about different approaches to treat asthma as of a certain date. Of course, he may have come up with 75 different approaches.

In PubMed, when you find a certain article, particularly if it's an important article, then you can access the references to that article. It says, "If you like this article, here are some other articles that are related to this article." That shows the power of PubMed. One of the dangers that exists from something like PubMed is that if you're writing an article, it's very easy to quickly come up with a list of references to some article. You may not have read all of the articles, but you can quote them as if you read them. PubMed would be a good tool for students to use who are plagiarists, because it makes you look smarter than you really are.

Another thing that was important about Stage 3 was that in some cases, we found very important articles that were published prior to the year 2000. While we wanted to limit ourselves to articles that were published after January 1, 2000, we found some important articles that were written before 2000, particularly in the area of case management and concurrent utilization review, which are older interventions

As you could imagine, PubMed is very sensitive to the MESH terms that were used. It's the responsibility of the PubMed staff at the National Library of Medicine to put in the MESH terms or the key words. We cannot claim that the literature search is comprehensive, because it's such a rapidly changing and growing field. I'd say that this is the best we can do in this time period.

Estimating the value of the interventions is difficult. We will go into the some of the methods in Paper 6. This is a very heterogeneous field when you look at different evaluations for the financial outcomes. The size and the methods that were used were very different, and how long they followed patients varied a lot. The time frame of the study could affect the generalizability of the results because sometimes a study was done in 1990 and then another study was done in 2003, and in the era of managed care, that is light years away. Some studies addressed specific diseases, such as asthma or congestive heart failure, whereas others focused on a particular intervention, such as preauthorization review or disease management.

The bottom line, unfortunately, is that estimating the value of interventions is difficult. We definitely found cases where we were convinced and the evidence was overwhelming that the disease management produced cost savings. In specialty care management, there were a large number of articles, but we didn't find any

definitive results that showed savings or non-savings. For population health management, which is the most recent sort of intervention, the admission rates declined 3 to 6 percent. These are sort of gross statements.

I recently did some work for some Wall Street firms. As part of that, they had what they call "investors conferences." Two of the speakers included the CEOs of Aetna Health Plan and UnitedHealthcare. I got to ask the question, "Do you think that disease management produces cost savings?" I was impressed because both of these people answered, "If you do it stupidly, it's going to cost you a lot of money. But if you do it very wisely, then it will save you money." I thought that was the best answer, and I think that the same is true with a lot of these other interventions here. If they are done carefully, then you can produce savings. If you set up these interventions without looking at the economic aspect of it, you're going to end up wasting a lot of money.

Chart 1 shows the results of the 106 different articles. You will notice that most of the articles involved disease management or case management. Population management is probably a little too new in order to have results, but I would expect more of those coming down the pike. Concurrent review and preauthorization/utilization review are like ancient history, so I don't think you're going to see more research in that area. I urge you to go to the Web site, because it involves greater detail on the articles that we reviewed.

Chart 1

Conclusions of the Literature

Intervention	Total Number of Studies	Major Findings
Preauthorization/ Utilization Review	9	Early studies show admission and bed-day reductions from UR in the range of 10% to 15%. Recent International studies of data not subject to managed care show considerable opportunity for utilization reduction. Early gains were not maintained as medical management models changed; there is also evidence of increased outpatient utilization due to inpatient UR. More recently these reductions are in the range of 2% to 3%; savings are estimated at between $25 and $74 per member per year; we estimate ROI of 4.60 based on reported intervention cost of $16/member for this study.
Concurrent Review	5	Early gains due to Concurrent Review were not maintained as medical practice patterns changed. Current evidence that Concurrent Review can reduce bed-days by 2% to 3%. One study in a hospital setting showed ROI of 0.9 (savings < cost of review).
Case Management	22	Reported results are variable (depending on target condition and program). Evidence exists of clinical improvement and reduction in utilization due to CM, particularly for heart disease. A survey of CM financial outcomes for Diabetes found no valid studies. ROIs in the range of 1.37 to 3.74 reported.
Specialty Case Management	5	Relatively few studies. Prevalence of members with target conditions makes them a poor candidate for randomized control trials. Evidence shows support for financial outcomes in mental health and some high-cost diseases, such as Renal Diseases.
Demand Management	6	Evidence exists that Demand Management reduces unnecessary physician and ER visits. Financial results indicate a return of between 1.37 to 3.86 to 1.0.
Population Management	7	Evidence reported of dollar savings within population wide programs. One study reported an ROI of 5.0 to 1.0. Studies of programs to intervene within entire chronic condition sub-populations report measureable pmpm savings.
Disease Management	52	For one population (multi-disease) program that reported pmpm savings, gross savings are estimated around $1.45 pmpm. For programs that report ROI, the range is 1.2 to 6.4. Highest savings are reported for heart diseases. Moderate savings are reported in diabetes and mixed results (in some cases no savings) for Asthma. A recent study using a randomized control showed no discernible savings.
TOTAL	106	

Now Ian is going to talk about the economics of care management.

MR. DUNCAN: To add to what Henry said, we summarized 106 of about 2,500 articles. We didn't read 2,500 articles (the Yale students did), but we did read the abstracts from almost all of those, which is a massive undertaking. Even to read in great detail, which we did, 106 different articles and then summarize heterogeneous, completely different articles into a paper that hopefully looks consistent, coherent and is able to balance the results of 50 different studies in disease management against each other is a very big undertaking.

Disease management is something that interests me a lot. If you look at the kinds of results that are out there in the non-peer-reviewed literature, people are publishing results that say that returns on investment (ROIs) are anywhere between about one to one, to somewhere around six to one. You see almost nothing about savings and ROI in the peer-reviewed literature. Michael Cousins, who's speaking tomorrow in the Care Management Symposium, has a paper published in one of the journals in which he actually quotes per-month-per-member (PMPM) savings numbers. That's very unusual; mostly you see an ROI quoted in the literature. I myself prefer to see PMPMs like Michael has done, and I think that we'd all be better off if we could get people to publish that way.

Briefly, in the time that remains, I want to talk about the economics of care

management. The industry, particularly the disease management industry, likes to talk about ROI. Most of those 50 articles on disease management published ROI statistics, meaning the ratio between savings and the cost of the intervention. The problem with the statistic, however, is that you can never figure out if a high ROI results from a high numerator or a low denominator or both, or whether the denominators are even comparable between different studies. It becomes absolutely impossible to make comparisons between different studies.

One of the things that has recently cropped up—I hadn't addressed this much before and only recently started to think about it—is that different vendors in the marketplace charge differently for similar programs. Some vendors will charge on a PMPM basis (number of heads in the population times a PMPM charge). Other vendors will charge according to the disease category of people who are being managed. Obviously over time in the latter case, you get a difference in mix in the types of diseases for which you're being charged. So not only do you have all the other things going on in that denominator, but you also have a change in mix between conditions. The denominator is changing, not because the price is changing, but simply because the mix is changing, making comparisons yet more difficult.

And there's always the old standby of random fluctuation. How much of high savings results is simply random fluctuation?

One result that I think is important but with which I am really struggling to express is the issue of scale. The disease management industry, largely vendor-led, has got to a certain scale, and programs have a certain size. By this I mean that there's sort of an accepted price per member in a commercial population that the industry and purchasers have agreed is about the right price to charge, and so they charge it. If you happen to have one million members in your health plan and the price that they charge you is $1 per member per month, your program is going to cost you $12 million annually. The average ROI, if you think about the range that people are generally quoting, is between 2 and 3. Let's say that it's about 2.5. That means that if the program is costing $12 million, then you have to save about $30 million. So instead of thinking about whether the right scale and scope of a program is $12 million or $3 million or $5 million, what happens is that the vendors are in hot pursuit of trying to prove that they've saved $30 million, because there is a level of cost and a level of price that everybody sort of thinks is about right in the industry.

There are some key drivers of the economic model. The one that I think is very important and needs more work in the industry is that of prevalence. Think about this. What is a chronic disease? Can people in this room even agree on what is a chronic disease? How do you identify and define people with a chronic disease? But what are the chronic diseases? There's not even a good agreement in the industry yet about what chronic diseases should be taken into consideration. You get a lot of arguments in the industry about what prevalence is. Some people make very inflated claims for prevalence of chronic diseases within their populations.

Obviously, the higher prevalence is, the more opportunity there is for savings. But in fact, in my experience, and I think there's some agreement, the major five chronic diseases—the two breathing diseases of asthma and pulmonary disease, heart disease, heart failure, and diabetes—within a commercial population amount to between 5 and 7 percent on average. If you see numbers that are around 10 percent or 15 percent, which you do from time to time, I think people are exaggerating and may be doing that because they want to be able to prove that they have higher savings or potential for higher savings.

I have a very simple example that I think points out the issues here. Think about a 30,000-life group, employees and dependents. A moment ago I said that chronic prevalence is probably in the range of about 5 to 7 percent. That would mean that, in total, a 30,000-life group would only have about 1,500 to 2,000 people with chronic conditions. Now you can dispute backwards and forwards what is a high-risk person with chronic condition, but let's just say for the sake of illustration that we agree that it's about 20 percent of the population. This means that we've gone from 30,000 total down to approximately 300 or 400 people who are considered high-risk, and therefore potentially worth offering a relatively expensive program to. Let's also assume that 60 percent of these people are “reachable.” One of the biggest problems in health plans, as those of you who work in health plans know, is that you don't have good telephone contact information on 30 percent of your population. You can't reach them at any time. With caller IDs, that number gets bigger. Then some percentage of the population, 5 to 10 percent, that you invite to be in a program is going to say no. So 40 percent of the people being unreachable and unenrollable in your program isn't out of the question at all in the average health plan. In two minutes we’ve gone from 30,000 total eligible members down to about 180 high-risk people that we can actually enroll in a program.

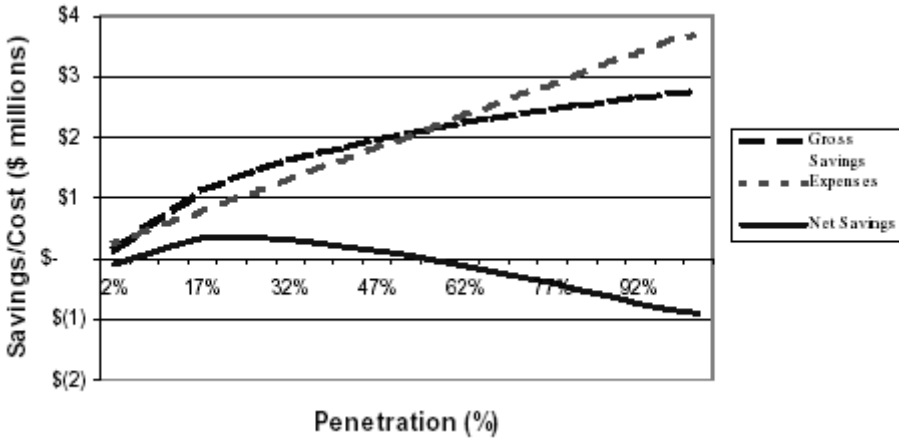
In a commercial population, the average admission rate on those people is going to be less than one per year. Let's say it's about 0.65. I don't have the math here and I can't do it in my head. Factor in also that the whole point of a disease management program is to change these people's behavior and keep them out of the hospital. Let's say that we're very good at this program and we can change behavior on 25 percent of the enrolled people. Maybe we'll be lucky that way and we can save between 30 and 40 admissions per year. That's 30 to 40 admissions on a 30,000-life population. If you assume an admission costs $8,000, your potential savings are about $200,000 to $300,000. If you convert that back into PMPM terms, it's about 60 to 90 cents PMPM. With that kind of economics it's very hard to justify a high price on a PMPM basis, or even a price like the $1.00 PMPM that I quoted earlier for a disease management program unless you can work very well. The one thing that disease management companies are very good at doing is influencing some of those variables. They are very good at finding the high-risk people, they're very good at the direct marketing aspects of reaching out to those people and finding them and they're very good at enrolling them. Those are the things that the disease management companies do well. If you really want to improve the economic performance of a program, that's where a company can help

you. But the economics are very marginal; you have to be very careful about the economics.

As you expand the size of a target population, the risk of that population is going to drop in the chronic population. You're going to have a very small number of people who have very high risk, and lots of people who are relatively low risk. I define "risk" as the probability of having a high-risk event such as an admission. As you drill down into a population and expand the reach of a program, the event probability declines. If you think about a program, yes, any disease management program is going to have a certain fixed cost, but largely it's an average cost kind of program. As shown in Chart 2, your cost rises more-or-less in a straight line the more people you get in, whereas your opportunity declines quite rapidly. Put those two things (cost and savings) together and you can figure out the potential gain from a program, which is the difference between potential savings and the rising cost of the program. Net savings will peak at some point, because there is a point at which it costs more to manage someone than you can save in avoided services. What point that is, I don't know for sure. It will depend on a whole bunch of factors, including prevalence, the shape of the risk curve and your cost structure. It's something to think about as you put a program together.

Chart 2

DM Program Savings/Cost



Contrast this with what I think tends to happen in the industry, which is that there's an accepted budget that everybody thinks is often determined by market norms as the right budget for a program (let's say a dollar per eligible life per month), irrespective of the opportunity in a population. Nobody looks at the opportunity in the population and constructs a program that way.

We think savings PMPM is the right thing to publish rather than ROI, for reasons

that I talked about a moment ago. We strongly urge the industry to increase the amount of disclosure. Having read through 106 papers on this and similar topics, I can tell you that the level of disclosure is totally inadequate. It's almost impossible to tell enough detail behind what's in a population, what the interventions were, how the people were identified and so forth, to really be able to assess and compare results of different studies. The amount of disclosure has to increase.

I want to talk a little about evaluating different savings methodologies (Paper 5). What makes for a good savings estimate? There wasn't a lot out there in the literature on evaluating evaluation methodologies, so we came up with our own set of criteria. Number one is that there should be a reference population. Outcomes measurement requires a reference population because without it, there's really no way of comparing against what would have happened. Statistics should be calculated consistently. This is fairly obvious, but it isn't always done by vendors. Something that we suggest, which isn't generally done in the industry, is that there should be some measure of appropriate measurement. In other words, measure only what the intervention is designed to manage. The intent of that is to keep potential confounding factors out of the evaluation.

The last criterion is the more actuarial concept of exposure. When you go through actuarial exams, they teach you about the concept of exposure and about measuring life years exposed or life months exposed, so we sort of grow up knowing about this. But in studies at which I've looked, there's much sloppier treatment of data, particularly on the denominator side, the exposure, the people whose heads are being counted and where those people are being counted. It gets as bad as analyzing the exposed population separately from the claims, and then dividing the one by the other, rather than making sure that there's always a person being measured in a particular month to whom you attach that claim. You get some anomalies and some distortion as a result of the failure to control exposure.

In order to evaluate different designs, again, we had to come up with some criteria. As you see in Chart 3, we tried to figure out ways to evaluate different designs. One way is, do they have scientific validity and rigor? A second one is, are they methods with which people are familiar? Sometimes somebody may come out with a new method, but you're generally not familiar with it, so you may not be as comfortable with the results. Also, is it a design that's replicable? Is it something that you can take the data yourself or take your own population, replicate and find similar results within your own population?

We've also summarized in Chart 3 how the method is often applied in practice. Finally, there are other miscellaneous issues.

Chart 3

Evaluating Savings Methodologies

	Method Type	Method	Validity/ Scientific Rigor	Famil-iarity	Replicability/ Auditability	Application	Evaluation of Methodology	Other issues
1	Control Group Methods	Randomized control	High	High	Difficult to replicate and audit; need another randomized group.	Requires untouched, randomized, control group. Metric in the intervention group is compared with the same metric in the control group, and the difference is assigned to the effect of the intervention.	"Gold Standard" method, although requires demonstration of equivalence. Need for incurred claims results in delays in evaluations.	Practical to implement and avoids adjustment issues, although requires sufficient number of members. Viewed by health plans as difficult to implement and potentially unethical. Randomization must occur at the population level if results are to be applied to the population.
2		Temporal (Historical) control	High	High	Replicable and auditable	Requires population drawn according to identical rules from two periods. Metric from the intervention period is compared with the same metric from the Baseline period, adjusted with trend. Requires adjustment of the comparison population to be equivalent to the intervention population.	Becoming the most widespread methodology in the industry. Need for incurred claims results in delays in evaluations.	Implicit assumption that regression to the mean is uniformly distributed in the Baseline and Intervention periods, and that a robust trend estimate is available.
3		Geographic or product line controls	High/Medium	High/ Moderate	Replicable and auditable	Requires population drawn according to identical rules from two different groups (e.g. geographies). Metric from the intervention period is compared with the same metric from the control, adjusted for all appropriate risk-factor differences.	Not widely used.	Sometimes difficult to adjust for the many risk factors that affect a population and its utilization (see Paper 2).
4		"Patient as their own control"	Low	High	Replicable and auditable	Patients are identified pre-intervention and then followed post-intervention. Pre-intervention metric is compared with post-intervention metric.	Widely used, but regression to the mean issues are causing purchasers to re-evaluate (see Paper 2).	Theoretically possible to correct for the effect of regression, but no method has yet been developed to do so.
5		Participant vs. Non-participant	Low	High	Replicable and auditable	Patients are invited to enroll in a program. Those who choose to enroll are subject to treatment; those who choose not to enroll form the control group.	Widely used, but selection bias causes this methodology to be highly suspect.	Theoretically possible to correct for the effect of selection bias, the effect of a member's "willingness to change" is unmeasurable.

	Method Type	Method	Validity/ Scientific Rigor	Famil-iarity	Replicability/ Auditability	Application	Evaluation of Methodology	Other issues
6	Non-Control Group Methods	Services Avoided (also called pre-Intent/post-Intent)	Moderate	High	May be difficult to replicate; auditable.	Record intent of different patients; track for a period of time to determine actual outcome, and assign a dollar value to the avoided event (adjusted for alternative treatment, if any).	Frequently used for small, highly-specialized programs (such as case management).	Two issues: participant bias (participants who are more likely to change their minds seek information and support) and evaluation and recording of intent is subjective.
7		Clinical improvement methods	Moderate	Moderate	Difficult to replicate; difficult to assemble comparable clinical trial data.	Measure clinical improvement and estimate financial savings using a model based on the difference in cost of well-managed and other patients.	Useful for small volume studies and when a result is required more quickly than data-based evaluations.	Requires review of the significant literature on clinical improvement, and a method for projecting financial from clinical improvement. To our knowledge there is no comparative study of results of clinical improvement and other methods.
8	Statistical Methods	Regression-discontinuity	Unknown	Low	Replicable and auditable	A regression line is fitted on the relationship between Year 1 and Year 2 costs in a population; Year 1 and Year 2 costs for the intervention group are then fitted and compared. A "shift" in the regression line indicates that the intervention has had an effect.	Highly-regarded as a theoretical method in the scientific literature, but we are not aware of a specific practical DM application.	To be determined.
9		Time-series	Low	Low	Replicable and auditable	Extension of the Adjusted historical control methodology to multiple periods.	Not widely used in commercial evaluations.	The effect of changes in risk-factors (often reflected in variations in Trend) is compounded over a period of years, making it very difficult to control this calculation.
10		Benchmark	Low	Low	Replicable; difficult to assemble valid comparison data	Metric in the intervention group is compared with the same metric in another population. The difference is assigned to the effect of the intervention and savings are estimated accordingly.	Occasionally encountered in commercial applications.	Comparison populations are unlikely to be described in sufficient detail to determine their degree of comparability (or the extent to which adjustment is required).

This chart may be difficult to read, so I'll go through some of it. There's a classification scheme that we've used here, which we came up with ourselves. We divided the different methodologies that are used in the literature into three types. The first of them is those that have control groups. The second is those that do not have control groups. The third is what we call "statistical methods." I'll explain those in a moment.

There are different types of control groups that are used. The one that everybody prefers is the randomized control method. It's rarely used in commercial applications, but it is quite frequently encountered in small-scale academic studies, which are primarily what we find in the literature. There are various kinds of non-randomized control groups that are used. One of them is temporal, or historical. There are studies out there where people use geographically distinct groups and compare the results with adjustments. There's a method that we call "patient as their own control," but which is also sometimes called "simple pre/post," where you measure the patient before the intervention and then the same patient after the intervention. Finally, you find studies (less so now, but older studies) in which the experience of participants is compared with that of non-participants.

These methodologies range, in terms of their validity, from high to low through that ranking. The participant versus non-participant suffers from the most potential bias. The least potential bias can be found in a randomized control method. Some of the ones in the middle, like the historical method or the geographic method, with the right corrections, can give you valid results.

There are some non-control methods that are frequently used. One of the most frequently encountered of these is what we call the "services avoided" method, which is often encountered in case management outcome studies. For these kinds of studies, the nurse records what's going to happen before an intervention and the nurse records what happens after the intervention. When the nurse has offered an alternative treatment or proved something alternative, the nurse simply records the savings. It's almost entirely a subjective methodology. It's not one that people like very much, but for interventions like case management, nobody has yet been able to come up with a better alternative.

Another methodology that we see from time to time is the clinical improvement methodology. It measures a clinical metric. For example, I've been able to get a patient who has a heart condition to take a beta blocker for a year. I look through the medical literature, and the medical literature tells me that taking a beta blocker improves something—whatever my metric is—by X percent. I assign X percent, find a value for that and assign that as a potential savings to the intervention.

Finally, there are three statistical methods that we talk about here. Regression discontinuity is one of them. That's a methodology that has a great deal of promise, but has not yet, to my knowledge, been applied successfully, at least in commercial settings, within disease management. There's a potential for time series. People talk about this, but it's difficult enough to do a one-period adjustment for the historical method, and if you think about extending that out for several years, it becomes even more difficult. The third statistical method people talk about is benchmarks. I can't think of a study offhand that actually uses external benchmarks. I think that this is something that will remain in the realm of theory. As I said a moment ago, generally insufficient information is disclosed about any study for you to be able to make the kinds of adjustments that you'd need to make if you compared your own data against a benchmark.

That's our very quick summary of the different methodologies. I encourage you to read those because it's a lot of detail if you're interested. In Paper 6 after the break we'll talk about the actual methodology that we used for some of the studies that we've done on data.

DR. DOVE: I want to make one comment that I think is important in this field. This is the role of Wall Street on a lot of what we're talking about. These disease management firms and the managed care organizations are very frequently pressured by Wall Street to produce results. This means that they want quick results, and they're not willing to wait two or three years before we can do the randomized clinical trial that Ian talked about. That makes all of this evaluation work more difficult.

MR. MARTIN E. STAEHLIN: I love coming to these sessions and learning about actuarial theory and then turning it into actuarial practice that gets people excited. I wonder if you found anything written about disease management or care

management specific to a client or specific to a population of people. After 35 years in exclusively health care, I think that there's a pooling point that's right for each case. Now you have to have enough people to be credible, but let's use your 30,000 people. So instead of trying to say, "Well, 200,000 would be the right pooling point for a large population," I think cases that big have individual pooling points that are appropriate, which implies that there's an underlying usage of health care. If you believe that (even if you don't), maybe there's a care management profile that's right for each case. If you go further, maybe there's a disease management profile, say a case that is prone to certain types of cancer or a population that's prone to diabetes. If that was true, then you could go in and do care profiles.

I'll put your session together with a session I already saw on predictive modeling, which said that 6 percent of the people consumed 58 percent of health care in a year. It also said that those 6 percent of the people consumed 23 percent last year and are predicted to consume 24 percent next year. That model implies that there's always a group of people that are getting really sick in one year. If you unwind that, the way you would maybe save money is to find those 34 percent of the people that are going to get sick and stop them from getting sick, which would require the employer to make them change their behavior, as opposed to saying, "Well, it would be really good for us long term if you would change your behavior." If you change what they pay for health care, now we're going to get into "my body chemistry makes me prone to these things, as opposed to my behavior." But let's set that aside for a second. I don't know if you found anything that would say that you should be able to find disease management programs for a specific employer. Instead of hiring a vendor to come in and say, "Here's my program," The vendor would say, "I'm going to design *for you* for your problem." I know that you said time series is difficult, but I've also seen a lot of case studies that say, "Certain employer groups are prone to these problems."

DR. DOVE: I haven't seen anything in the literature, but I would think that if you look at the historical fights that occurred between physicians and hospitals in managed care organizations, you would have a revolution with what you're thinking about with respect to more aggressive kinds of interventions. I do think that there will be new kinds of interventions over the next five years. We're starting to see more interest in consumer-driven health care. We're seeing more experiments with different kinds of deductibles in co-insurance and tiering of hospitals. We're seeing more interest in pay-for-performance, which is a different idea. Who has ever heard about pay-for-performance for the patient? We talked about pay-for-performance for a physician or for a hospital; maybe what you're thinking of is some kind of pay-for-performance for the individual. But I haven't seen anything like what you've described.

MR. STAEHLIN: We're saying that risk measures are really getting in. So just like Medicare can attach a risk score to 39 million people, you could attach individual risk scores to your employees. You don't have to say to them that they have to see the doctor three times a year. But if your risk score is over some benchmark, you

know what? You're wasting our money, so you ought to pay more.

MR. DUNCAN: I'm not sure that this answers your question, but I've said this before in these forums about disease management. As somebody who used to run a company and who has actually put programs in place and run call centers full of nurses, you're relying on these programs to change people's behavior to comply and conform with certain treatments. I always felt that I was fighting this battle with one arm tied behind my back, because the disease management company is often at risk and the health plan is at risk, but the one component of this puzzle that's not at risk is the patient. Until we move health plans more in the direction of consumer-driven health, where people have to take on more financial responsibility for their own actions, we'll continue to be fighting with one hand tied behind our backs.

MS. SUSAN DENISE MAXWELL: I have a two-part question. I'm looking at a couple of perspectives here. First of all, I would like to thank you for undertaking this. This has been a subject of real interest for me, and I've spent a fair amount of time looking at various articles.

All of the information that I've heard and these ROI studies have focused on savings within health-care costs. My first question is, has anyone looked at financial impacts outside of health-care cost savings? For example, has anyone looked at reduced absenteeism in the workplace? Those kinds of areas are outside the medical care cost.
My employer's largest product line is stop loss (I work for Highmark Life & Casualty Group). My second question is, has anyone looked at cost savings above certain dollar amounts? Say a $50,000 or $100,000 specific deductible?

MR. DOVE: I'll answer the first question and then I'll turn it over for Ian for the second question. We didn't look at absenteeism or short-term disability costs. It was hard enough to estimate the cost savings for the various interventions. Frequently you had a reduction in emergency room visits or a reduction in hospitalizations. Sometimes you would get data on charges, and then you would have to convert charges to cost and so forth. We were not able to go to that level of detail for absenteeism, but that's certainly a very important issue.

MR. DUNCAN: Yes, that is even more difficult than what we were looking at. I think the literature is early in that area; I haven't seen anything much.

As to your question about stop loss, I think chronic people tend to be in that middle band. There are a lot of chronic people who use relatively little health care resources, and there are quite a lot of them who have fairly high costs. But for the most part, you're talking about people whose costs probably average between $5,000 and $10,000 a year. The stop-loss cases, depending on your attachment point, tend to be people who have things other than your average diabetes or your average annual maintenance for heart conditions and so forth. That's about the

extent of what I've seen. I have seen nothing specific on really high-cost claims.

MR. KENNY W. KAN: It has been my observation that many published studies on disease management do not appear to back out the impact of mean reversion. If you do so, basically your ROI will be much less. Is there a theoretically pure approach that you would recommend that would address this?

MR. DOVE: Come to the next session. Ian will be talking about how to overcome the difficulties or the fallacy of regression to the mean. That is a very important topic that we'll address.