RECORD, Volume 30, No. 1*

Spring Meeting, Anaheim, CA
May 19-21, 2004

Session 111OF
Measurement of ROI/Effectiveness of Medical Management Processes

Track: Health

Moderator: Ian G. Duncan

Panelists: Ian G. Duncan
Robert N. Parke
Jaan Sidorov†

Summary: Actuaries are increasingly being called upon to advise management about the value of medical management programs and interventions. Outcomes measurement raises many issues, some of which actuaries are familiar with and some that are new.

MR. IAN G. DUNCAN: I'm Ian Duncan. I'm a consulting actuary at Lotter Actuarial Partners in New York. We have three speakers today. I'm going to talk a little bit about some of the research that we are doing on medical management outcomes measurement sponsored by the SOA's Health Section.

Rob Parke, who is with Milliman in New York and is chairman of the Academy's Working Group on Disease Management (DM) Outcomes, is going to talk about some of the work done by his group putting together what will probably ultimately be a practice guideline for actuaries working in this area. He'll also share with us some recently published Milliman research into some of the DM outcomes that they have recently published.

Our third speaker is Jaan Sidorov. Jaan is a physician with Geisinger Health Plan in Danville, Pa., and he is going to be speaking to us about some of the work being done in the DM industry. Jaan is chairman of the Disease Management Association of America's (DMAA's) Research and Quality Committee. He's going to talk to us

† Jaan Sidorov, MD, not a member of the sponsoring organizations, is the medical director and care coordinator at Geisinger Health Plan, Danville, Pa.

about some of the work being done and some of DMAA's publications to help us in this area. This is a research-focused or a presentation of research results-focused session. I'll let the other speakers introduce themselves in more detail, but this gives you some idea of the reach of this panel.

As I said a moment ago, we're going to be talking about some of the research currently underway. The SOA's Health Section is sponsoring a study of the application of different measurement methodologies in DM and medical management more generally, and the working title of this is "Actuarial Issues in Disease Management Outcomes Measurement." I'll be talking more about that in a moment. Rob will talk about the Academy's work, and then we'll hear from Jaan about work being done more generally in the industry.

By way of background, Chart 1 gives me the opportunity to put in a funny but important quote by P.J. O'Rourke: "If you think health care is expensive now, wait until the government provides it for free." As our keynote speaker said on Wednesday, remember that while we all complain and moan about how costs are increasing and how dysfunctional the health-care financing system may be, the system does deliver enormous benefits, tremendous quality and tremendous technological advances. Those of you who watched Dr. Hughes' presentation will remember that curve that trends quickly down from the late 1960s in terms of deaths from heart attacks. It is important to remember some of the benefits that we get from the system.

If you work in DM or care management more generally, one of the things that is true if you study the literature is that the clinical results are impressive. There's a study that I'll recommend to those of you who are interested in it done by a professor named Scott Weingarten published in *The Lancet* in 2002, "Interventions used in disease management programmes for patients with chronic illness—which ones work?" S. Weingarten, Henning, JM, Badamgarav, E, Knight, K, Hasselblad, V, Gano, A, and Ofman, J. Meta-analysis of published reports. BMJ 2002; 325.

He did what is referred to as a meta-analysis. In other words he collected all the studies and summarized their results. He looked at more than 100 published studies of DM outcomes and concluded overwhelmingly from the studies that DM improved clinical outcomes. It improved hemoglobin A1c scores and lowered cholesterol values. However you want to measure these things, DM conclusively improved clinical outcomes.

A similar meta-analysis exists, but it's not published yet. A man named David Krause did it, who I think is at Marquette University in Milwaukee. His conclusion was that if you looked at the published financial studies of DM outcomes, the results are inconclusive. There are some that seem to show savings and some that seem not to show a savings. There's basically inconclusive evidence of savings. There seems to be an anomaly that on the clinical side we see good results. Where are the good financial results? Why don't they come through on the financial side?

Part of it may have to do with the way the programs are structured because it seems fairly clear if you look at the implementation of DM programs that frequently at least early ones were implemented to produce beneficial clinical outcomes rather than beneficial financial outcomes.

Some of it may be that you get the results that you structure the program to produce, which are clinical outcomes, and that if you want to get good financial outcomes, you might have to structure a program slightly differently. Another reason for it may have to do with the way that we measure these things. We'll be talking more about that during the session. The president of one of the large DM vendors said to me recently, "I know that we don't always show these savings in the financial results, but I know that we do good. "

One of the papers—the SOA's Health Section study that I referred to a moment ago—is going to come out as a series of papers. The first two are introductory and background papers and deal with the industry and the issues. They are going through a second stage of review currently, and I'm hoping that they will be posted on the SOA Web site (www.soa.org) when the peer review is complete. Those of you who are interested can access them there. We had originally intended in the early papers to include a literature review similar to the study made by David Krause of financial outcomes, and we found it extremely difficult to find published, peer-reviewed, sound articles that gave any financial outcomes at all.

We had hoped to make this part of the initial publication, but it's taken so much longer, and I've had to go out and hire a couple of graduate students at Yale to help because it's proving to be so difficult to find good, peer-reviewed financial outcome studies. That may come out a little bit later. There's some other work that I'll refer to in the next few minutes. What we've seen so far is similar to Krause's results: that there seems to be weaker support for financial savings than for clinical improvement.

The bottom line on the published literature in this area seems to be clinical results are good, but the jury may still be out on the link between clinical and financial outcomes, which may still need to be established. The financial jury is out. Why is the financial jury out? One of the things that drives me nuts in this area is everybody talks about ROI rather than per month per member (PMPM) savings. I will have more to say on this point in a moment.

We've all been at the point where we think some of the claims made by the vendors are perhaps unrealistic. There is no generally accepted accounting principle for these financial measurements. Another thing that drives me nuts working with clients is poor reconciliation controls. The type of thing that you'd expect to see as an actuary—you start with some numbers that balance to some published financial outcome and work toward a result—is something that isn't done often in the industry. There's a lack of understanding, I believe, and attention to the things that drive financial outcomes as opposed to clinical outcomes. The bottom line on all the

financial discussions comes back every time to a typical statement, which is, "If you saved me all this money, why does my trend keep increasing?"

When you join the DM industry, as I did several years ago, there are three words that all begin with R that you learn on day one and repeat. Whether you know what they mean or not doesn't seem to matter, but people repeat these words. To try to standardize definitions, DMAA has a working group right now coming out with a dictionary or handbook of DM terms to try to standardize the definitions of some of these commonly used terms. The first of the three R's is regression to the mean. Everybody talks about regression to the mean, but it's not always clear that everyone is talking about the same thing. Rob Parke will be showing real data on this in a moment.

The second is ROI. The third is reconciliation, which in the DM industry means what I, as an old underwriter, would call an annual accounting or a savings calculation. It does not mean what I think is important and what I referred to a moment ago, which is the process of reconciling your results back to some source data that you had some confidence in. I think this is one of the areas where we have to do more work in the industry if our results are going to have credibility.

I promised I'd say another word about ROI, so bear with me here for a moment. ROI is a ratio, and I don't think it's the best financial measure for health plans. It's defined as the sum of savings divided by the sum of costs. If you want to compare programs or compare outcomes, remember that you're comparing a ratio. The first question is what's in the numerator? How do you define your savings? That's a difficult enough problem on its own before you add a denominator to it. Adding the denominator increases the complexity.

There are a couple of limitations with the measure. It fails to identify the source or the bottom-line results. Do you get higher ROI because you have high savings or low costs? Do people define costs the same way? Do you include all the costs, such as internal costs and external costs, or do you include just the vendor cost? What do you include in that denominator?

I think one of the more interesting challenges in terms of understanding is if somebody came to you and said, “I have this investment that can return 300 percent on your money," which is three times ROI, you'd say, "That's a good investment, but tell me more." If I put some more money into it and got a 200 percent return but was able to triple my absolute savings, which should I invest in? The industry talks about high ROI as though it is a good result in and of itself but doesn't ask whether it has an optimal program or an optimal design, and has squeezed the maximum TOTAL savings from the program.

Think about an average health plan. What's a target return on capital nowadays? When I worked for Aetna it used to be 15 percent post-tax. At the margin that meant you should be earning 115 percent on investment, or 1.15 ROI, and yet here

are programs that are returning 300 percent. They're clearly suboptimal. It makes comparisons between programs, between vendors and between health plans impossible to do. I think we need to do some more work in determining a good metric to measure these programs by and publish results on.

I want to talk briefly about some of the preliminary results from the work that we are doing. I headlined this part of the presentation "Games People Play," but I don't mean that people do play games, at least not intentionally, but that the way that you measure the outcomes can be gamed to some degree. I'll touch on four aspects. The first of them is the identification of the chronic population itself. The second is how you measure trend. The third is how you measure enrollment. The fourth is the measurement methodology itself.

Regarding chronic identification, all people are not created equally or identified equally when it comes to chronic disease. What do I mean by that?

Jaan could probably tell us a way that you could objectively and unquestionably identify people as chronic from their medical records, from tests or from some clinical markers, but those of us who do program management and program measurement have to rely on administrative claims data. When it comes to administrative claims data we're all familiar with the problems, and what we get for identification when we use claims data is absolutely not an objective identification of chronic people.

One reason is that it's hard to set a single set of criteria that everyone will agree with as implying that somebody is chronic or not. The first set of criteria is what I've called a narrow identification, which basically says the individual has had an inpatient claim with primary diagnosis for a chronic disease and some type of face-to-face medical service on the outpatient side (Table 1).

The second set of criteria is a broad definition, which says that somewhere on an individual's medical record, there is an inpatient claim with a diagnosis for a chronic disease, or other medical services of any type, including tests.

The third set of criteria is based on the old narrow set plus outpatient prescription drugs. You get a different set of chronic disease prevalence depending on where you draw your identification. This matters when it comes to measuring outcomes because this is the population that you're going to select from for both management and measurement purposes.

Table 1

Prevalence of 5 Chronic Conditions

	Narrow	Broad	Rx
Medicare	24.4%	32.8%	30.8%
Commercial	4.7%	6.3%	6.6%

LAP Client data; duplicates (multiple conditions) removed.

One of the things that those of us who work in this area have learned as we've gone through this is that you don't always have to measure the outcomes of the same population that you manage. You may be content to measure a subset of a broader population. What does this mean? It means that you can set a set of criteria that are sensitive to identify potential candidates for a program, but when it comes to measuring your outcomes you might want to get a set of criteria that are much more specific. You trade off sensitivity against specificity, and you might set your borders in different places depending on the need for that chronic identification.

What happens when you apply the narrow, broad and Rx-based definitions to a particular chronic disease? Table 2 happens to represent diabetes. A population identified in year one as being diabetic, using my narrow set of criteria, which basically are the hospital-based primary diagnosis criteria, identify only 30 percent of what I might identify by applying all criteria to identify that population. If I add in my broad definition, I can increase that number substantially, and then by going to look for drug claims such as insulin, I can bump the number up yet again. You find different groups of people through applying different criteria.

Table 2

		Narrow	+Broad	+Rx	TOTAL
	Year 1	30.4%	47.7%	21.9%	100.0%
Year 2	Narrow	23.1%	18.1%	6.7%	47.9%
	+Broad	4.4%	21.2%	5.8%	31.4%
	+Rx	1.2%	4.3%	7.9%	13.4%
	Identified	28.7%	43.6%	20.4%	92.6%
	Not Identified	1.7%	4.2%	1.5%	7.4%
	TOTAL	30.4%	47.7%	21.9%	100.0%

LAP Client data; duplicates (comorbidities) removed.

Look at what happens in year two (Table 3)

Table 3

		Narrow	+Broad	+Rx	TOTAL
	Year 1				
Year 2	Narrow	75.9%			
	+Broad		85.5%		
	+Rx			92.6%	
	Not Identified	24.1%	14.5%	7.4%	
	TOTAL	100.0%	100.0%	100.0%	100.0%

LAP Client data; duplicates (comorbidities) removed.

If I apply my narrow criteria from year one to the same population in year two, not everybody is going to qualify again in year two, and it is important to understand what the implications of this are. Here's a simpler view of the same data. If I go narrow in year one and narrow in year two, only 75 percent of my population qualifies again in year two. That means that if I follow the same population from year one to year two, I'm carrying 25 people out of 100 who somehow or other

don't meet my claims criteria and, by definition, will be low-claiming people. They must be because they don't meet claims criteria for the disease.

If I'm going to measure the outcomes of those people, I have to be aware of the fact that unless I select out people who do not meet the criteria in year two, I'm going to be carrying forward some statistical false positives from year one to year two who may be depressing the financial outcomes and therefore the trend in the chronic population.

If I expand my definition to the broadest possible set of criteria, I still don't get a 100 percent match from year one to year two. I still get a degree of statistical false positives of 7 percent, but I'm also now dealing with a broader, less risky population of people with probably lower claims and probably — Jaan can probably help us with this — the broadest definition might find a lot of people who are clinical false positives rather than the statistical false positives that I found first. I don't have a good answer to this. I think it's something that people should be aware of when they're watching or measuring the outcomes. How you define your population is going to impact your results.

The second issue I said I'd talk about briefly is trend. This is something that's actuarial. The reason a trend matters is that probably the most popular way of measuring outcomes is to take a baseline population, apply a trend and compare it with a population that's been intervened on. Savings is equal to trended baseline less actual.

In the DM industry I don't believe there's the same grasp of the complexities of trend that there is in the actuarial profession, so people will say, for example, "Trend is 15 percent. I know that because I read it in the *Wall Street Journal*. I'll apply 15 percent trend to my baseline population." The result will of course be a high savings number. We don't need to belabor the point because we all understand the complexities of trend, but clearly this is something where actuaries have a lot to offer the industry. A lot of the debate going on right now is whether to use a total trend, a chronic trend or nonchronic trend.

One of the issues that I think impacts trend and is worth thinking about is what you do with the point at which you identify somebody as chronic. It matters whether you identify people prospectively or retrospectively. Let me put up some more examples. Obviously, if you're identifying people from claims, there's going to be a point at which somebody moves from being nonchronic to chronic as identified through claims.

When I'm calculating my trend and find somebody halfway through a period, should I go back and take that person out of my historic population or include him in my chronic population but retrospectively? *When* you move that person from one group to another significantly impacts the measured trend, and here are some examples from client numbers (Table 4).

Table 4

	Trend
Chronic trend, chronic members Included prospectively:	-1.1%
Chronic trend, chronic members Included retroactively:	7.2%
Nonchronic trend, chronic members Excluded prospectively:	2.4%
Nonchronic trend, chronic members Excluded retroactively:	6.1%

Don't take away the specifics, but get a feel for the size and the direction of the impact.

The first line says I'm measuring chronic trend, that is, trend from a chronic population year one to year two where chronic members are included prospectively, meaning from the point at which they were identified through claims. Here I get a small negative trend.

If, on the other hand, I go back and put those people retroactively into my baseline population and measure their trend, I get a different number. The same directional effect is seen on the nonchronic trends. Again, I don't have a good answer to this one, but be aware that there's a lot of movement in trend numbers, and *when* you identify somebody, be it prospectively or retrospectively, can have a significant impact on your trend and therefore on your savings calculation.

The third issue is that of continuous enrollment. For this I'm going to acknowledge some work done by Michael Cousins, who's a colleague of ours on the DMAA Research and Quality Committee. What the numbers in Chart 2 show is something that won't be a surprise to many of you. They basically show how many people out of a population of 100 identified at the beginning of a period survive through three months, six months, nine months and so forth. That's simple.

But look what happens if you measure savings in a DM population the same way but apply different requirements for continuous enrollment on those populations (Chart 3). At the extreme the highest savings numbers exist when you require that people be enrolled continuously for 24 months. If you cut back to six months, you get a different number but much lower and so forth. Intuitively this makes sense because if you require that somebody be enrolled for 24 months continuously, you're obviously going to exclude those people who died during the period, and we all know that people who die have the highest claims. Again, how you define these things can have a significant impact on the answer.

Finally, let's discuss methodologies. There's been a lot of debate about what a reasonable methodology is to use for doing these types of studies. Jaan's going to talk more about this. The DMAA has published a white paper on some of this with some principles (available at www.dmaa.org). One of those principles that's important is that if you're going to do a study of savings or outcomes, you should have some reference population to compare your intervention population against, and I talked a moment ago about a basic idea where you use a baseline and trend it forward. That's a reference population to compare with the actual. Savings are defined as the difference between the trended baseline and the actual.

One methodology that's popular is what I call "patient is their own control," and what that means is I identify a population of people who are diseased in year one and follow them for 24 months after intervention compared with preintervention, so pre-/post, but I'm only following people who are identified in year one. This is a classic example of regression to the mean. Rob's going to put up a lot more information about regression to the mean at the individual level. I'm putting up a simple Table 5 to remind you what can happen in a population.

Table 5

$' 000		Projection Period Cost Distribution			
Moderate Historic Cost	Baseline Period Average Cost	$0 - $2	$2 - $25	$25+	Projection Period Average Cost
$2 - $25	$5,658	$ 668 55%	$ 6,599 40%	$47,811 5%	$ 5,398

Source: LAP data; 210 continuously enrolled members of an HMO; 1998 Baseline; 1999 projection

It is important to understand the difference between regression at the individual and population levels. Regression can occur at the individual level and yet, because of the random nature of these changes in individuals (some increasing, some decreasing) not be observable at the aggregate, population level. These numbers show that regression is also observable at the population level, despite the offsetting changes in individuals.

When you identify a population, as we have here, in the baseline, the average cost of this chronic population is $5,658. This is a population, by the way, that we followed for 24 months. There was no intervention of any type other than some fairly simple preauthorization. There's no case management and no DM. We are watching a population that is not subject to any intervention. What happens to the population? Looked at in year two, the projection period average cost is $5,398. That's approximately 4.5 percent or 5 percent lower, and this is a population with

no intervention at all. Nobody's touched them, and their costs have come down, despite underlying trend in utilization and cost. You have to be careful when you follow individuals or populations in this way that you are able to identify the natural progression of their costs as opposed to the effects of a DM company or an intervention.

Those are four things for you to keep in mind in the future. I'll remind you that some of these results will be coming out over the next few months on the SOA's Web site. Those of you who are interested in it should check the site from time to time. Jaan is going to speak next and is going to talk about some of the work being done in the industry.

DR. JAAN SIDOROV: Thank you. As everybody knows, I'm a physician. I know little about actuarial things . I am a stranger in a strange land. However, I beg your indulgence, and in exchange for that I promise to have only one actuary joke in my presentation.

One of the reasons why I agreed to speak today was not only because of my delight in being able to speak to any forum at any time about DM, but also because of my ability to always learn something new in attending conferences like this. That's my main motivation, and I'm not only interested in hearing Ian's and Rob's presentations, but I'm also going to be interested in the questions..

I'm going to try to talk about three things. The first is a physician's viewpoint and what many of my colleagues on my side of the aisle are thinking about quality, insurance and managed care. I'll talk a little bit about the experience of one person—that's me—in dabbling in DM, and I'm going to show you some typical data that were recorded in the typical peer review literature that forms the coin of the realm in DM, and then I'll close with some discussion about the Disease Management Association of America, what it's doing and some of the help that you can give to help to keep the rest of the industry honest.

As a physician, I'm going to tell you that any of you at any time can make an appointment at Geisinger Medical Center to see me, and when you see me you'll be given a 20-minute appointment. That's the platform. You and I, once the door closes, will enter the special land, a magical place called the doctor-patient relationship. The fact is that the doctor-patient relationship, which has been in place for a long time and which is financed by the health-care system, is dysfunctional.

There's an increasing body of literature that's showing that it doesn't work well and that if you come and see me and rely on me to be your doctor, there are all types of errors that could happen in terms of missing values in charting, in terms of followup, in abnormalities of lab results and in not being charted correctly. I may not fill out a prescription correctly. There may be all types of drug interactions. There's an increasing realization through the Institute of Medicine and other bodies

that the health care industry can do a better job when it comes to quality. Everybody agrees with that.

The managed care industry in particular has done a lot of work. I think there are a lot of reasons for this and for what its motivations are, but the managed care industry has done a lot of work in trying to promote quality through the National Committee for Quality Assurance (NCQA), through publications and through the Health Plan Employer Data and Information Set (HEDIS). I'm sure all of you are familiar with this. The reason why the managed care industry can dabble in this is because from a physician standpoint and from health services research, the managed care claims data sets have a lot of valuable information. We have demographics. We know age. We know gender. We know zip code. We know employer status. There are diagnoses that are available through ICD-9 and CPT coding, and we know how much is being spent on things.

It may not be known, but in the peer reviewed medical literature it turns out that a lot of work that's been done in finding out what types of interventions and systems of care promote quality has been done by managed care researchers using managed care dollars through foundations. The best example is the Kaiser Health System, which has done some good work in coming up with interventions that make physicians do a better job of screening for sexually transmitted diseases, promoting mammography and promoting colon cancer screening, and these are examples of that. However if you do a meta-analysis, and this is one (there was another one recently published in the *Annals of Internal Medicine*), it turns out that managed care versus any other type of health insurance has not been conclusively proven to promote quality.

For 10 years a lot of physicians, particularly during the '90s, thought that managed care by virtue of the fact that it's managed care would be the platform to increase quality in the United States. If you compare head-to-head clinical trials, and there aren't that many, and if you pool population data and look at what's happened to people who are in a fee-for-service environment versus a managed care environment, the quality of health care—the event or that 20-minute office visit—isn't necessarily going to result in better colon cancer screening, better mammography screening, better treatment of your multiple sclerosis, better control of your hypertension or better control of your diabetes. It's not there. It's one big, red herring.

Here's my joke. The actuaries apparently have known this for a while. You put on your hat, do whatever it is that you do and realize that the types of metrics that I use to measure quality in terms of diabetes, in hypertension care, in colon cancer screening, in mammography and in all those other things don't seem to do too much to the medical loss ratio at all. A lot of physicians out there have also intuitively realized, and the evidence is growing, that the type of insurance that you have when you walk into that 20-minute office visit isn't going to make any difference in quality. Does quality make a difference at all?

My physician community is doing a lot of work in all these areas trying to promote these things to show that it does make a difference, and it will ultimately noodle through into the medical loss ratio and hopefully do something about the cost of health care inflation.

Leaving the financials and insurance aside, there's some additional bad news. There was a study put out by Leatherman et al., who performed a series of case analyses visiting these particular institutions (Chart 4). Visits were done on good programs such as diabetes management, use of group visits and tobacco cessation. They're all listed here. Two of these in particular, the smoking cessation programs and diabetes management, were in-house DM programs, and it turns out that in every particular instance, the business case was terrible.

The institutions lost money. They lost money in trying to promote these types of things. If an institution develops a diabetes management program and hires nurses to do all the things that nurses are supposed to do to promote the clinical guidelines promoted by the American Diabetes Association, it loses money.

The reasons listed are there's a failure to pay for quality while paying for defects, consumers can't perceive quality, a consumer-payer disconnect exists, there's uneven clinician access to relevant information and there are displacements in time and place. None of them should be surprising to any of us. The business case is difficult for quality.

But somewhere, someplace, somehow, somebody has to be ultimately saving money. If somebody stops smoking, 20 years later there is a payoff somewhere. If people with diabetes do a better job of checking their feet, the decreased rate of amputation somewhere down the line is going to result in a salutary financial effect somewhere. It's that there's no ROI within one year that the chief financial officer (CFO) is going to be able to hang his hat on.

Going back to insurance and not being able to let go of this, now we're entering the realm of how DM people think. Usually, insurance risk is segmented by domain. There's a separate bucket of money for primary care. I'm a primary care physician. Our bucket of money is relatively small. There's money for specialty care. There's money for durable medical equipment.

How each of those dollar flows out of those buckets done through capitation or fee-for-service varies, but my interpretation of what the DM industry is saying is that perhaps it's possible to segment risk by condition and that instead of thinking about this in terms of primary care, specialty care or in-house care, let's talk about the cost of care for patients with diabetes and take patients with chronic illness such as diabetes and congestive heart failure (CHF) and devote buckets of money to that and see how much money at the end of the year is left. If there's money at the end of the year, you have that pot at the end of the rainbow called ROI. How does DM do it?

What's its proposal or thinking in terms of trying to save money? This is what it does, basically. If you take the cap off DM and look inside the jar, this is what you'll find: a big emphasis on clinical guidelines to reduce variation, promotion of literature-proven care strategies, intense patient education and self-care and intense case management, usually all performed by nonphysician personnel.

We have that 20-minute office visit going on, but before and after you see me, either a nurse is seeing you or a pharmacist or a case manager will see you, contact you, e-mail you, telephone you and in some way intervene and tell you that there's a guideline out there. Case management will do all types of patient education so that you understand as much about your disease as possible. That's what it does.

The interesting thing is that if you go to a Web site such as the American Healthways Web site, there's a lot of talk about collaborating with physicians. As an architect of some internal DM programs, we generally leave the physicians alone. They're busy enough. They don't like it when you call them up and tell them how to practice medicine. We leave that 15- or 20-minute office visit unchanged. The other thing that is going on in the DM industry is that there's significant investment in information technology, registries and gigabyte or terabyte servers that hold all the claims data, all the clinical data, all the quality-of-life data and all the satisfaction data, segmented by patient and by condition, and that's what they're using.

This is one example of one institution's foray into the area of DM. This is a study (Chart 5) that appeared in *Diabetes Care*. The authors, and I'm one of them, were involved in a study where the managed care organization randomly selected patients with diabetes for a HEDIS study. It turned out that about half of these patients had voluntarily opted in after being approached by a nurse to participate in DM. The other half did not. It was a quasi-experimental, self-selected, nonrandomized, clinical control study using a population that otherwise appeared completely equivalent.

Both groups, as far as we could tell, were similar in terms of age and gender, the prevalence of the pharmacy benefit and enrollment in the plan. Both groups were similar. By having a nurse call these patients, promote clinical guidelines, do the case management and do the patient education, the HEDIS measures all moved in the right direction, which makes sense. In addition to seeing your physician as often as you want, a nurse also is coaching you and prompting you to do the right things. That's basically how this works.

The interesting thing is that over the course of a one-year period, these people consistently had a lower claims pattern to the tune of about $104 per patient per month, which, carried out times 12, compared to the cost of the program. I told my CFO that on the basis of the study, it looks as though we have a good ROI, and he bought it.

There's another study underway by American Healthways. It was engaged by Blue Cross/Blue Shield Minnesota to develop DM programs in a number of areas. The areas are listed in Table 6.

Table 6

	% Changes: $PMPM		% Changes: ER Rate		% Changes: Admission Rate	
Program	**Study**	**Reference**	**Study**	**Reference**	**Study**	**Reference**
Asthma	4%	17%	-17%	-1%	-13%	12%
Cardiac	-10%	23%	-25%	3%	-47%	6%
COPD	-10%	32%	-23%	6%	-28%	40%
Diabetes	12%	19%	-12%	2%	-1%	2%
Impact	10%	21%	-20%	-3%	-6%	7%
Core Programs	2%	20%	-18%	1%	-25%	7%
Overall Results	6%	20%	-19%	-1%	-16%	7%

SOURCE: BLUE CROSS AND BLUE SHIELD OF MINNESOTA

Patients who bought the commercial insurance were the subjects of the DM program. Patients who were in TPA-employed groups were not in the DM program because the TPAs didn't buy it. Both groups were cared for in the same network and had essentially the same benefits. After American Healthways moved to the care of these populations, it's purporting to find that its program was associated with a consistent decrease in claims expense, and this is a report from *Managed Care* magazine. Interestingly enough, this hasn't appeared in the peer review literature, and a lot of the data are simply percentage reductions. I don't know what exactly the numbers are that the company is purporting to show, but there you see the return of at least $2.90 for every dollar invested.

They also claim a projected 2 percent to 3 percent reduction in the total, fully insured, commercial health care expenditure rate. That's what the company said. It said that it's shifted to Hopkins for additional analysis. We'll see. The thing is, though, you still aren't too sure about this stuff. I can't say I blame you. It's difficult enough taking past claims experience plus trends and coming up with future experience in rates, and it does not compute in terms of quality for chronic diseases, which is one of the reasons why it's so important that the SOA and people like Ian are thinking long and hard about this.

Here's the mission: advocacy for DM, of course. We're called the Disease Management Association of America. We're interested in promoting quality and standardization of outcomes. What is the content? What are the outcome measures? What are the definitions? By editing the Dictionary of DM Terminology

for DMAA, Ian is doing an important job in making sure that we all mean the same thing when we talk about ROI or covered population and use all the other terminologies out there that are bandied about in the DM industry

The tasks are to define measurement methods, including financial, economic, clinical and quality; to promote preferred methods; to create the definitions; to create an industry snapshot; and to standardize satisfaction.

The Quality and Research Committee is insisting among its member organizations that the DM industry also needs to get more serious about reporting its outcomes. It is possible in many clinical settings, including the DM space, for outcomes to be measured with a higher degree of rigor instead of using the usual pre-/post design with all the bias that Ian talked about.

It is possible to carry out some randomization, particularly in rollouts, case control and convenience cohorts. This is all out there in the peer review medical literature. We're asking that the industry itself get involved in trying to at least incorporate some of these methodologies when it's reporting outcomes in the peer review literature. The Governmental Affairs Committee is helping the Centers for Medicare & Medicaid Services (CMS) to work on the pilot program project, Section 721 of the recently passed act. We're also working with the Congressional Budget Office and its efforts of predictive modeling.

Chart 6 is a segment out of the white paper that's been posted on the Web site, and I'll talk to you a little bit more about that because I'm coming up on the end. This is a graph that describes scientific rigor on the y-axis versus real-world feasibility for DM evaluation on the x-axis. As you can see, it's unlikely that the DM industry will ever get to the point where it's conducting randomized control trials. That is not going to be done because there is no funding, and there's no market interest in doing that sort of thing. However, there are some other high-rigor methodologies available, and that's what we're trying to promote.

We have a white paper posted on the Disease Management Association of America Web site at www.dmaa.org. You are all welcome to log onto that Web site and take a look at that paper. Here's the quick lowdown. The author is the Quality and Research Committee, and Ian helped us with this. There is no standard approach and no single approach to evaluating outcomes in the DM industry. There are some preferred approaches and high scientific rigor approaches, such as using an equivalent control group, that are possible in the DM industry.

If there is no control group, the white paper suggests that at least the entire chronic population, however that's defined, be studied with careful descriptions of how that population is defined. The other big problem in much of the literature on DM is that there's insufficient transparency. It's difficult to discern exactly what the nurses, the pharmacists or the case management people did to whom and how they did it. For the actuaries to be able to discern whether or not there's any real

savings, you need to be able to understand that. Is it just a program that mails leaflets to patients hoping to see a 3 percent to 5 percent regression to the mean that is supposedly the savings, or is there some meat to the program?

I encourage you to look at the paper. Also be aware of the coming work on the definitions project so that when we use the various terms that are being bandied about in the literature we all mean what we say, and we say what we mean.

MR. ROBERT N. PARKE: I'm Rob Parke. I'm an actuary with Milliman in New York. I'm currently the chairman of the Disease Management Work Group for the Academy, and that's the reason that I'm here. I also wanted to show you some of the research that Milliman has been doing recently regarding some of the issues. I'm not going to be giving you any answers to these issues. At this stage I think where we are with some of the analysis and where actuaries can contribute is that we need to start quantifying and discussing the issues before we come up with any solutions. I think we quickly realized at the Academy that if we tried to define anything, we'd be here until 2015 before we get anything out into the public arena, and our intention is to do that.

What I want to talk about briefly is ROI at a high level. Because we're talking about ROI, I'm going to touch on cost, but I don't think the cost element of a DM program is where we add much value. I'll then specifically touch on some of the issues that need to be considered when discussing program savings, and I know there's some disagreement about that. In my mind, when you're looking at DM programs, two issues are in some way unique to DM programs, unlike many of the other analyses that we do, such as trend and the difference in provider reimbursement, which we often have to deal with. That's not to say that these issues aren't confounding, difficult to deal with, but they're not unique.

I want to talk specifically about regression to the mean and selection bias. The reason we're talking about ROI, as Ian has said, is that escalating health care costs are causing employers and insurers to reassess the value and effectiveness of their medical management programs, and with that I throw in DM, and they're looking at DM to prevent major disease events and hopefully to reduce the need for in-house costs of medical care.

I would like to add something to what Jaan said, which is bear in mind that sometimes it's cheaper if people die. Because they're living longer doesn't mean from a cost perspective you are saving money. This is something that some of the national health systems think about. In the U.K., for example, you can't get dialysis on the national health service after age 65, and there are good reasons for that. There are interesting moral dilemmas and debates about that, but they do need to be factored in in some way into this equation.

MR. PARKE: At a high-level ROI equals total program savings divided by total costs. There's a huge debate about how to effectively define this. I think Ian

forwarded to me recently what the DMAA is putting out. It's a paper on ROI to get the discussion going, and I'm sure that he would be happy to send that to anybody who wants it. I don't want to get too much into that but to recognize that that's something.

When we talk about cost, again I'm not focusing in on cost because I don't think that's an area that we can add too much debate on because this is much more of an accounting issue, but many times you need to take into account all of the costs. If any of you have done these analyses where you're trying to estimate the ROI or the impact of a DM program, often the in-house costs are not taken into account, and they're a key component in my mind to some of these costs.

Savings is the area where I think we can start adding something to the equation. By way of putting it into perspective, Alison Johnson, one of the consultants at Milliman, issued a research report some years ago of a survey of the different methods that were out there to measure most of the costs. These are the three that came back, The first one was a comparison of pre-enrollment medical expenses with postenrollment medical expenses, yet this is straightforward because during the course of managing a member's disease, programs often approve or deny payment for services based on protocols for managing disease savings, and the savings are basically calculated by comparing requested services with approved services. That's one method that's used.

The second method is the medical expenses of enrollees in a DM program are compared with the medical expenses of a group of people who aren't enrolled in the program but have the same chronic disease as the enrollees. This one is not typically used because there are difficulties getting around that.

The last one is the one that most people are using, and I'm not judging these in any way whatsoever. The last one is the one that most DM vendors use these days to estimate the savings attributable to the DM program, and this is where the total health care cost of all the enrollees for the year or some time period prior to enrollment in the DM program are compared with the same enrollees' health-care costs during subsequent periods after enrollment. As you can well imagine, there are significant, confounding analytical issues in making sure that you have a like-with-like comparison. That's where I think that we, as actuaries, can add something to this debate. These are things that we think about every day in the work that we are doing.

In terms of the confounding actuarial issues, the two issues that I mentioned earlier were regression to the mean and selection bias, and I want to share with you what I mean by regression to the mean. Basically what regression to the mean is saying is that human beings get better irrespective of a significant financial event, irrespective of the intervention of a DM program or irrespective of the intervention of any medical professional in many situations. That's what regression to the mean is at its simplest, and you can understand what that means. If costs are going down

when you're measuring pre- and postintervention, is it as a result of the typical way that a disease progresses or is it as a result of the intervention of the program that we're looking at?

To illustrate some of these issues, we at Milliman recently did some research that looks at the progression of various diseases. We were looking at heart failure (HF), coronary artery disease (CAD), diabetes and comorbid diseases on the Medicare population and the commercial population. For the Medicare population we used the MedScan databases which are '98 to '01 for active employees and dependents with drug benefits, and we excluded HMO and point–of-service with capitation from this analysis to remove any potential impact that contracting might have on the costs of the DM.

From the Medicaid side we used the 5 percent sample from '98 to '01, and again bear in mind that the results between commercial and Medicare are not directly comparable because the Medicaid data doesn't include prescription drugs. All of the costs were trended up to '01.

On the commercial side the trends were based on total PMPM increases for the population members identified with the disease state that were continuously enrolled during the four-year period that we were talking about. On the Medicaid side the trend estimates were taken from the '03 annual report of Medicare's board of trustees. Before I forget, you can download this report from our Web site. The title is "Insight into Two Analytical Challenges for Disease Management."

Bear in mind that this is what happens with a Medicare diabetes average claim cost over the period that we were talking about (see Chart 7). Again, when we were talking about a significant financial event, the significant financial event at the top is defined, and you can go into the research report and see that as a hospitalization or emergency room (ER) visit coded with a diagnosis of or related to the conditions we were talking about—HF, CAD, diabetes and comorbid diseases—and you can see that this is the impact on costs as the disease progresses. This is without the intervention of any DM program.

The number is $23,000 at the top, and that's dropping off to $2,800 at the bottom. There's a significant tailing off in the drugs, and you can see that it doesn't regress back to the pre-enrollment costs before that. That's for the Medicare diabetes, and we see a similar pattern when we're talking about CAD. Again it's going from $28,000 to $2,600. This is all in '01 dollars on the Medicare side.

We can see the same thing for the HF, and you can see a slightly different pattern. There seems to be a longer tail here in the HF. There's an 80 percent drop from the peak down to the Q5 costs for Medicare. For the comorbid disease, you can see that longer tail again. That's going from $28,000 to $4,800 in terms of these comorbid diseases.

You can see the same thing on the commercial side. I reiterate they're not directly comparable because the commercial includes prescription drugs where we obviously don't have the prescription drugs on the Medicare analysis, which tends to suggest that we can't identify people through the use of drugs for the condition.

We're probably looking at a slightly more serious population in the Medicare side. We've divided the data between the over 18-year-olds and the under 18-year-olds for the diabetes. Again it's going down from $17,000 to $3,000 for commercial diabetes. I'm not giving you any answers in terms of how to address these issues, but this is a major issue when you are trying to get an effective calculation or estimate of the impact of any DM program.

Commercial CAD and is almost exactly the same, going down from $28,000 to $3,000 in '01 dollars. The comorbid disease is in Chart 8. Remember that this is without the intervention of any DM program. This is looking at fee-for-service data.

The next issue that we covered in some of the research is selection bias. The selection bias basically says that members participating in a DM program may be different from those not participating in the program, and that has implications for how you calculate the impact of the program. This may result in significant utilization and cost differences between those enrolled and those not enrolled when you're making these cost comparisons. Using the same data as we've been looking at, we made an attempt to estimate or illustrate the potential impact of the selection bias. We split the population into three different population groups.

The terminally ill are the potentially high cost population. These are people who are typically excluded from any DM program. We then looked at the compliant population. These are people who are motivated to look after themselves, and you'll see what I mean by that. We then looked at everybody else. For the various disease states and the various populations, we looked at the difference in costs, which will have implications.

The definition that we used for the terminally ill or high-cost cases were discharge disposition is death or hospice, any outpatient hospice care and individuals with date of death. It's more detailed in that research report.

Working with our clinical consultants, we came up with this definition of compliant for diabetes, which is that they had any recommended diabetic-specific study or exam in the year that showed up in the administrative claims data that we're looking at and one face-to-face encounter not in the ER.

For CAD we had a different definition of compliant. For HF, with or without CAD, we defined compliant as one face-to-face ambulatory encounter not in the ER; one prescription for Ace inhibitor (not Medicare); for commercial, no readmission or ER visit within 30 days for HF; and for Medicare, no more than one readmission or ER visit for HF in the quarter of a significant event. Neither was everybody else.

As you can see, Table 7 shows the results of the selection bias

Table 7

Disease State	Population	Terminally Ill/ Potential High Cost	Compliant	Neither
Diabetes	Commercial	$11,806	$4,561	$1,768
	Medicare	$42,784	$43,809	$16,859
HF	Commercial	$30,532	$19,637	$11,457
	Medicare	$44,823	$50,678	$21,433
CAD	Commercial	$12,448	$14,813	$5,870
	Medicare	$36,760	$41,451	$14,680

In '01 dollar terms you can see that there's a significant difference in the costs of these programs, and this has major analytical issues when you're trying to estimate the true costs of your DM program. Those were by way of illustration the major unique, confounding, analytical issues of a DM program. There are a lot of other actuarial issues that need to be taken into account, such as provider reimbursement, benefit designs, claim adjudication and trend. I'm not going to cover any of them. These are issues that you and I deal with on an ongoing basis in most of the work that we do anyway, and that's not in any way to suggest that they aren't difficult issues to deal with. They are ones that are not particularly unique to where we're going on this.

To finish, I wanted to give you a background on the Disease Management Committee of the American Academy of Actuaries. The Disease Management Committee of the Academy is in the process of putting together a discussion document for exposure to the membership. We believe that we need to get something into the public arena if we want a seat at the table in terms of anything going on. As a first step we're not going to try and do anything amazing because we won't get any agreement on this, but we're trying to get into the public arena a discussion document that raises all of these issues.

It will be an articulation of potential issues. Ultimately we are hoping that it will lead to a practice note of these issues. The other thing that is of interest to the Academy is the Medicare Prescription Drug Improvement and Modernization Act that act that was passed last year. It provides for the phased-in development and evaluation and implementation of chronic care improvement programs. One of the things that is part of that developmental process in that act, is they're calling for "randomized control trials in the evaluation of quality improvement measures, beneficiary and provider satisfaction, health outcomes and financial outcomes."

We believe that we have something to add in this debate, and we would like a seat at the table. The Academy has a coordinating committee, which includes the heads

of all of the working groups that have input on where this new bill has impact or overlap, and we are coordinating those responses. That is one of the reasons why we think it is imperative to get something out into the public domain, so at least we have a seat at the table while these issues are being discussed.

One area of concern for some of us is that if we don't get involved in this, we risk losing it to the academics. One of the problems with some of the academic research is it lacks a business focus to some extent. I could have given you a definition of quality-adjusted life years, which seems to be the favorite of many of the academics in this area. It has its place in a discussion or assessment of DM programs, particularly when you're talking about a national health system or something like that, but from the point of view of employers, I don't think that it is of major importance, and I think we can add something to this debate.

MR. DUNCAN: Thanks, Rob. We will open for questions.

MR. BILL LANE: I've done work similar to what you've done in terms of your false positives and your regression to the mean. It wasn't for DM purposes, and, therefore, it's not the chronic conditions you did. I was looking at cancer. I found similar regression to the mean numbers where with constant dollars you could see drops of as much as 30 percent to 50 percent. I saw similar false positives that Ian was talking about, but one thing I did notice within the numbers because I had a database that was geographically dispersed was there was a significant difference in the false positive rate by geography, which probably relates to practice patterns, and a higher false positive is going to give you more regression to the mean. You'll have more people who truly weren't with that condition and therefore will drop down the average the next year. Otherwise I thought it was interesting information.

MR CHUCK FUHRER: I want to ask Mr. Parke whether on the regression to the mean analysis you did any more work than look at the mean for the people with the chronic conditions and then look at the mean in the following period because that's all you showed on the slide.

MR. PARKE: I'm not sure I understand your question.

MR. FUHRER: I was expecting to see a regression coefficient up there.

MR. PARKE: There's more detail in the report..

MR. DUNCAN: I'm intrigued by this compliant versus neither comparison in the numbers. Not having looked at these numbers before, it strikes me as counterintuitive in a way that people who are defined as neither, which I take to mean not compliant in some way, should have lower costs than people who are compliant. Isn't this clinically counterintuitive? Shouldn't the people who are compliant with everything we tell them to do be lower cost than the people who aren't?

MR. PARKE: They result in more office visits because they're compliant. They result in higher drug costs because they're compliant. This is a small time period that we're looking at.

MR. DUNCAN: Yes, but the differences here are amazing. They're factors of two or three.

MR. PARKE: Again, I don't think the intention of what we were trying to do was to say that that is a compliant population in the DM sense. It was an attempt to illustrate the potential impact of those issues. We were taken aback by the variation.

MR. DUNCAN: Yes. We have some data and some analysis that I think I've shown at these meetings before based on randomized controls that we ran in my prior company that show that people who enroll in these programs tend to be lower cost. The more compliant people are more likely to enroll.

MR. TOM SNOOK: Thank you for your three informative presentations. I was thinking about this calculation of ROI and specifically how you determine the savings. There's a key issue that, Jaan, you touched on briefly, but I'd like the panel to talk about a little bit more. It's the timeframe over which one measures savings, especially for certain things like smoking cessation, cholesterol or hypertension. The big returns may be 10, 15 or 20 years down the road, and it seems that on the one hand a true, fair measurement of the ROI would capture that on a present-value basis somehow, but then that logic also ignores the realities of health insurance, the employer base and a health plan that's spending money on DM today, which probably won't be covering that individual 10, 15 or 20 years down the road. Its competitor will or Medicare will. Could you talk about that a little bit?

DR. SIDOROV: I'll answer this from a clinical perspective. In the area of diabetes, much to our surprise, as well as to others in the industry, when the preliminary data started coming out, the return—however you define that—was seen within the course of a year, and intervening in patients' diabetes and getting better blood sugar control in month one resulted in fewer nondiabetes complications over the course of the next 12 months. I'm on thin ice here, I understand that, but it looked as though the claims for patients involved in our diabetes program went up in the first quarter and then went down, and that's been seen in other DM programs. Most of the cost savings that we reported in the manuscript in *Diabetes Care* appeared to be in the area of arthrosclerotic events at 12 months, not diabetes complications. The original hypothesis, the way you save money over the course of five years in diabetes care with less blindness, less renal failure and fewer amputations, has not been demonstrated in the DM literature because the DM industry hasn't been around that long. Rather, it's nondiabetes complication at 12 months that seemed to have dropped, and some bench research has shown that when patients have better blood sugar control, there are other salutary effects in

terms of what all those mediators and cells do at the level of the artery that seem to be not as twitchy and resulting in thrombosis, heart attacks, strokes and related conditions.

MR. DUNCAN: The reality, of course, is that for a program that starts on January 1, the CFO will be there on January 2 demanding to know what the ROI is.

MR. PETER HENDEE: I am surprised and frankly relieved that all of the focus on the cost difference is on the cost in the medical plan, and nobody considered productivity losses, absenteeism and all of that. I was wondering what the direction is on that. I've seen it before, and it's not there now. I'm relieved it's gone, but I'm wondering why.

MR. PARKE: Certainly from my point of view it's not gone. We just didn't highlight it in that presentation. I think that this is clearly something that employers think about, and it's part of the DM cell. I don't think it's gone. I think that people are trying to focus on some harder numbers at the moment because they've had some bad experiences in terms of being promised savings that in their minds have never materialized. I don't have a good answer for you about that, but it's something that needs to be factored in in some way or other.

DR. SIDOROV: To make it even more complicated, I attended a seminar by my colleague Pat Salber who's located here in California, and she pointed out that the metric for absenteeism and presenteeism is also subject to the same deficiency as savings calculations, namely the absence of a standard measurement methodology. No one knows how to measure absenteeism in the workplace because of how we all take vacations, sick days and personal days, and it's a difficult thing to count.

MR. PARKE: Can you define presenteeism for us?

DR. SIDOROV: Presenteeism is the phenomenon of having a sick person at work who isn't being as productive as he should be.

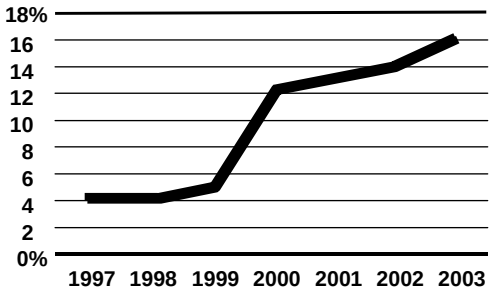
FROM THE FLOOR: Is that in the dictionary?

MR. DUNCAN: No, but that's a good thought. I should include it in the DMAA's Dictionary of DM Terminology.

Chart 1

Healthcare Trend has accelerated

"If you think Healthcare is expensive now, wait until the Government provides it for free" (P.J. O'Rourke)

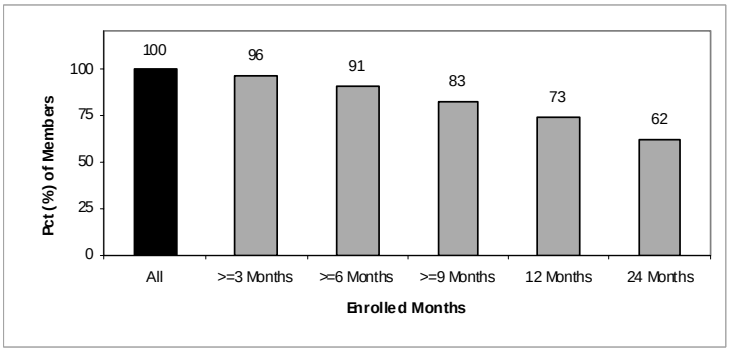


Source: Towers Perrin Health Care Cost Survey. 3

Chart 2

Continuous Enrollment

Some methodologies require 12 months of continuous enrollment in both Baseline and Intervention
Effect of Continuous Enrollment Criterion on number of members in Evaluation*



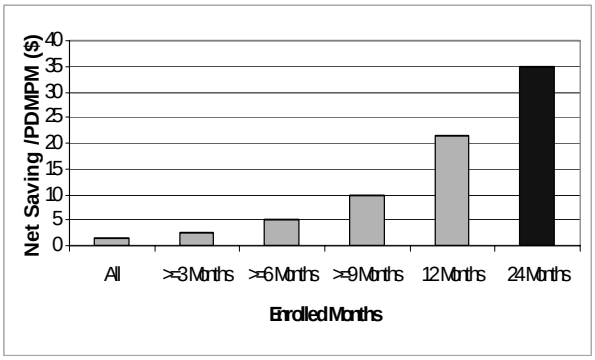
*** From Michael Cousins, PhD: "Evaluation Parameters and Continuous Enrollment Savings" Unpublished Study. Reproduced by permission of the author.**

15

Chart 3

Continuous Enrollment

Continuous Enrollment is inversely related to Savings*

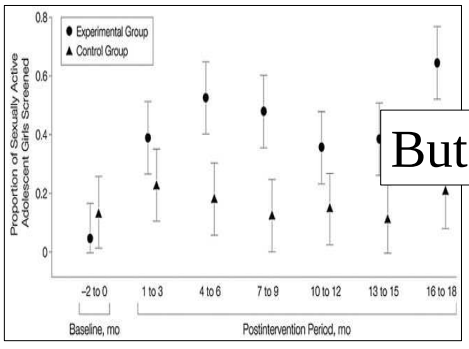


*** From Michael Cousins, PhD: "Evaluation Parameters and Continuous Enrollment Savings" Unpublished Study. Reproduced by permission of the author.**

16

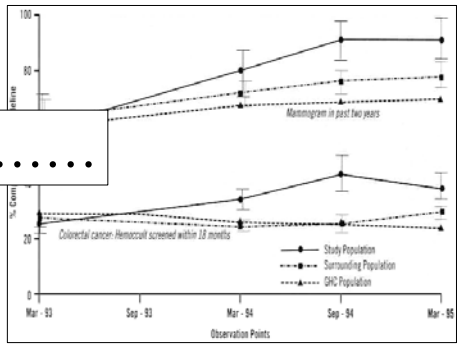
Chart 4

Reports in the peer review literature



Shafer et al: Effect of a clinic practice improvement intervention on Chlamydia screening among adolescent girls. *JAMA* 2002;288:2846

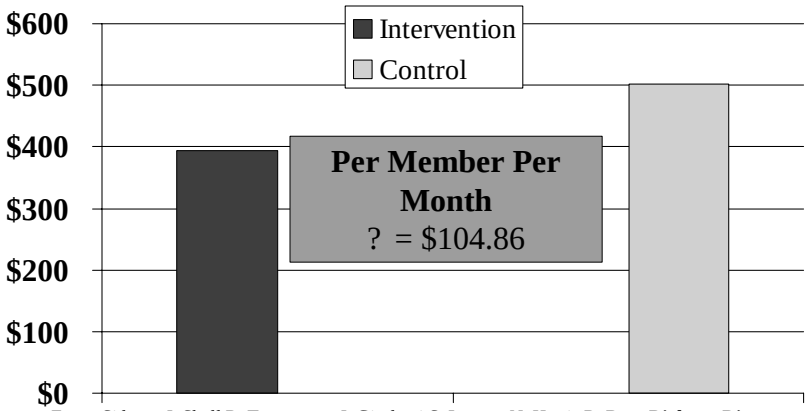
But……



Taplin et al: Putting population-based care into practice: real option or rhetoric? *J Amer Board of Fam Prac* 1998 11:116

Chart 5

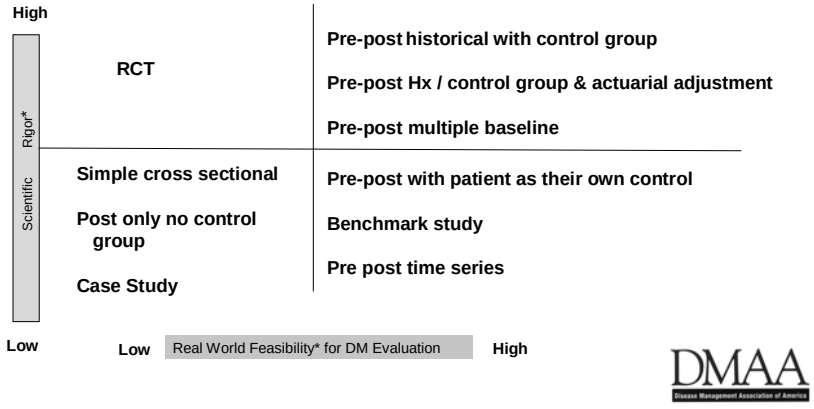
Diabetes Mellitus Disease Management Savings



From: Sidorov J, Shull R, Tomcavage J, Girolami S, Lawton N, Harris R: **Does Diabetes Disease Management Save Money and Improve Outcomes?** A Report of Simultaneous Short Term Savings and Quality Improvement Diabetes Care 2002; 25:684-689

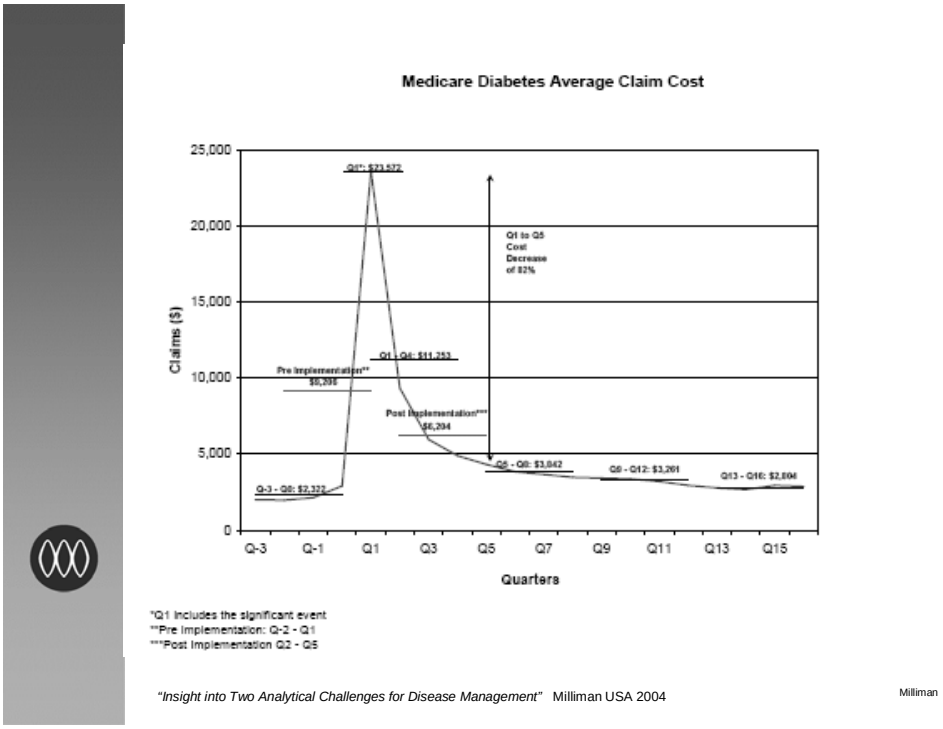
Chart 6

Trade-off Between Scientific Rigor & Real World Feasibility for DM Evaluation



DMAA
Disease Management Association of America

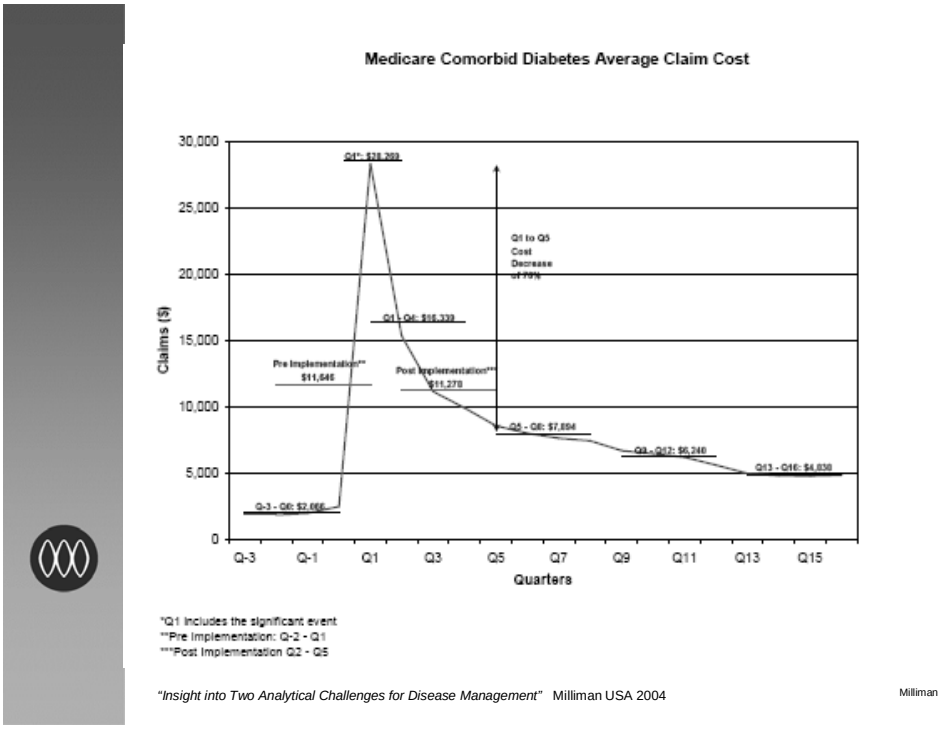
Chart 7



"Insight into Two Analytical Challenges for Disease Management" Milliman USA 2004

Milliman

Chart 8



"Insight into Two Analytical Challenges for Disease Management" Milliman USA 2004

Milliman