

RECORD, Volume 25, No. 2*

Seattle Spring Meeting
June 16–18, 1999

Session 73IF Profiling Physician Performance

Track: Health

Key Words: Health Maintenance Organizations, Management Information, Modeling Tools, Models, Providers

Moderator: JOEL C. HOFFMAN

Panelists: DAVID G. JOSEPHSON
DAVID L. TERRY, JR.
KELLY MYERS[†]

Recorder: JOHN C. VATAHA

Summary: Panelists discuss the methods HMOs employ to measure their providers' quality of care, outcomes, compliance with best practice guidelines, and cost control.

Panel discussion covers the following:

- *Measurement methods*
- *Implementation issues*
- *Success factors*
- *Reaction from physicians*
- *Improvement in quality and cost-effectiveness*
- *Steps HMOs can take to improve physician performance*

Mr. Joel C. Hoffman: I'm with Ernst & Young LLP. I'm sure this is a topic of interest to everybody here. I do a lot of work consulting to payers and providers, and profiling physician performance is of great interest to me. I am sure you have heard a lot of these phrases:

- "I haven't seen any data"
- "I don't understand the data"
- "The data are inaccurate"
- "The data are not timely"
- "The data are inconsistent"

*Copyright © 2000, Society of Actuaries

[†]Mr. Myers, not a member of the sponsoring organizations, is Vice President of Strategic Development at MEDai in Orlando, FL.

Note: The charts for this session are not available online. Please contact Linda Blatchford at lblatchford@soa.org or call 847/706-3507 for a hard copy.

- “The data are not representative”
- “My patients are sicker”
- “These aren’t my patients”

I do not want to call them excuses, but I am sure you have encountered some of these comments when you work with a managed care organization or a physician organization trying to understand the data you have put in front of them to help improve performance. I always like the last one, “These aren’t my patients,” as well as “It does not reflect their population” or “It is not risk-adjusted to the benchmarks.”

Again, I am sure you have come across these issues that physicians or providers have with the way we try to educate them on their performance levels and where we would like them to move.

Taking responsibility for profiling of providers is difficult. There is never a clear understanding of who is responsible. Usually, whoever is responsible is the one who is going to get the blame. It is really a cooperative effort and an interactive process that has to be ongoing. Why do we profile? We profile to identify and share best practices, to size the delivery system appropriately, to identify poor performers, to determine and adjudicate approaches to provider reimbursement and incentives, and to support disease management initiatives and internal pricing efforts. I’m sure you have all encountered these issues as well, but the main idea in the managed care industry today is about improving performance, doing better, and delivering higher quality care that is cost-effective. To do that you need to disseminate best practices, which means you need to provide data to your providers. You really need to understand when there is average performance. You need to educate the providers who are performing at that level to help them understand how to improve performance and how their peers have done better with the same patient populations. I always find some problem in terms of the data feedback. I think we will get there soon enough in a couple years when there will be a payer and a provider working together to get this iterative process going so that it is used effectively and productively. As of yet I have not seen a seamless process that everyone has adopted and believes in.

We talked about physician profiling being a cooperative effort. Everyone has to buy into the concept and the measures. Ideally, everyone would help develop the measures and use them for the common goal of delivering quality care. We talked about physician profiling being an iterative process. It requires all these pieces of information in order to happen. You should not forget compliance issues when considering profiling. Profiling also relates to education and service delivery and has to be timely. These are my keys to success, and hopefully our experts will elaborate on them.

Essentially, I think profiling has to be simple. It also has to be understandable and believable. That involves a lot of educational effort and a lot of development effort up-front. The need for timely and accurate feedback is usually among the first things that I hear from the provider regarding profiling. The data do not come to them in an accurate form. It is not timely enough to allow them to use the information to the benefit of the patient population they are covering. And, the issues that you have to overcome relate to whether your profile is an appropriate measure for the physician's population, considering risk adjustment. The physicians have to buy into it and believe in it or it is not going to work.

I want to introduce our presenters. When I was putting this session together I thought it would be interesting while looking at physician profiling if we had someone who represented a payer, someone who represented a provider, and an independent vendor who has some type of tool or capabilities that help us perform physician profiling. I think each of the panelists will have different viewpoints that will help us better understand this topic. Dave Josephson is going to lead off with the payer perspective. Dave is a vice president with Aetna U.S. Healthcare and one of their leaders in their strategic healthcare analysis unit. He has 20 years of experience in the industry, 13 of those in managed care. His responsibilities currently are medical cost analysis, medical management program evaluation, and performance measurement. Prior to his current position he was chief actuary at Aetna U.S. Healthcare.

Representing the provider side is Dave Terry from PhyCor, a physician practice management company. He is vice president of contracting and actuarial services and has 18 years experience in the managed care industry. His current responsibilities are pricing, reserve analysis, and contracting on behalf of PhyCor. Our final panelist is Kelly Meyers, who is representing the vendor perspective. Kelly is vice president of business development for Medical Artificial Intelligence (MEDai). MEDai offers medical artificial intelligence solutions to healthcare payers and providers and is currently focused on addressing the more difficult challenges facing our industry such as physician profiling.

Mr. David G. Josephson: As Joel mentioned, I am with Aetna U.S. Healthcare. I am with a subsidiary of Aetna called U.S. Quality Algorithms (USQA). Performance measurement is one of our major objectives. The three questions I want to address are: (1) What is it that USQA is measuring by way of provider performance? (2) How do we go about performing that measurement? (3) What do we do with the results? As far as what do we measure, for primary care physicians (PCPs) there are a number of different indicators. We measure member satisfaction. We send out hundreds of thousands of surveys each year, compile those results, and feed that information back to PCPs. We also measure how PCPs manage the care of members with selected diseases, including diabetes, asthma, and cardiac care. We also look at the appropriate utilization of

prescription drugs. Obviously prescription drug costs have been a very hot topic lately, so that is something of great interest to us. Among the other things that we measure for PCPs is over- and underutilization. Many of our PCPs are capitated. In a capitated environment we are particularly interested in underutilization, but we are also interested in tracking overutilization as well.

For specialists we have developed performance measurements for ten different specialties: including cardiology; urology; ears, nose, and throat; and gastroenterology. The top ten specialties make up approximately 80–90% of the specialty costs. Some of the things that we measure are member satisfaction, which primarily utilizes information we gather through surveys. We also have measures that are developed specifically for each of the specialties. For example, for obstetricians we measure their C-section rates.

For hospitals we look at inpatient care, particularly the average length of stay. We think that is a measure of the efficiency of the care that is provided in a hospital. We are also interested in the rate of adverse events such as readmission for the same diagnosis. We also look at infection rates and other adverse events that can come about when someone is in the hospital. We also measure C-section rates because maternity is a big cost item. When we are measuring hospitals we can drill down by the specialty area within that hospital. We can also look at the managing specialists of that care so that we can actually begin to pinpoint within a hospital where their performance may be deviating from an average. For pharmacies we look at the appropriate use of drugs for selected chronic diseases as well as member satisfaction and drug interactions.

How do we measure our providers? As you can imagine, information comes from disparate sources. Therefore, we have built a data warehouse that combines the information from medical claims, pharmacy providers, encounter information that we receive from our capitated arrangements, and member and provider data. We use primarily administrative data as well as survey results. We use administrative data because it is readily available and a cost-effective way to begin to build the performance measurement.

Examples of the desired types of data include patient ratings of care or satisfaction surveys, provider satisfaction surveys, chart reviews (to the extent that we use those for other purposes), outcome measures, member demographics, provider information, hospital information, medical claims, encounter data, pharmacy data, patient management, laboratory results, and premium data.

With this data we have built a member health profile database that identifies members with 1 or more of 65 or more diseases. Using primarily diagnosis and procedure types of information, we have been able to identify members with

these kinds of diseases and to develop prevalence rates. The particular disease that is in bold type on Table 1, congestive heart failure, suggests that, at least at one particular point in time, 1.2% of our total membership had congestive heart failure. Within our self-insured population it was 0.7%. In our Medicare population, as you might expect, it was over 9%. For our Medicaid business it was about 0.2%. For our commercial population about 1%.

TABLE 1
DISEASE PREVALENCE RATES
(ILLUSTRATIVE ONLY)

Disease Classification	Composite	Self-Insured	Medicare	Medicaid	Commercial
Asthma	7.0%	6.0%	4.2%	11.1%	7.0%
Atrial Fibrillation	0.5	0.3	4.9	0.0	0.4
Back Pain	7.7	7.5	16.4	1.8	7.6
Benign Prostatic Hypertrophy	1.0	0.7	10.6	0.0	0.8
Cerebrovascular Disease	1.0	0.6	9.9	0.1	0.8
Cholelithiasis	0.9	0.9	2.4	0.3	0.9
Chronic Obstructive Pulmonary Disease	0.8	0.6	6.4	0.1	0.7
Chronic Renal Failure	0.2	0.1	0.9	0.0	0.1
Congestive Heart Failure	1.2	0.7	9.3	0.2	1.0
Depression	3.1	2.3	3.2	1.7	3.2
Diabetes Mellitus	3.3	2.5	15.8	0.9	3.0
Diverticular Disease	0.8	0.6	6.4	0.0	0.6
Epilepsy	1.0	0.9	2.0	0.5	1.0
Hypertension	8.9	7.6	41.0	1.7	8.3
Hypothyroidism	1.7	1.2	4.5	0.4	1.7
Ischemic Heart Disease	3.2	2.4	23.3	0.3	2.7
Migrane and Other Headache	4.8	4.2	4.2	2.2	5.0
Osteoarthritis	1.5	1.3	10.2	0.1	1.3
Peripheral Vascular Disease	0.7	0.4	7.5	0.0	0.5

Once we have developed the health profile database we risk-stratify members into five severity levels. Severity levels are intended to be indicators of the propensity for an acute exacerbation with respect to that particular disease. Chart 1 is an example for congestive heart failure. As you can see, Severity 1, which is the lowest level, has 42% of the membership whereas the highest level, Severity 5, has about 5% of our members with congestive heart failure. The bar graph on Chart 1 shows the acute bed days associated with those different severity levels.

We use logistic regression models to project the expected measures. We then compare individual provider results to the overall Aetna U.S. Healthcare averages. Those overall averages are case-mix-adjusted for such things as age, gender, clinical service group, or diagnosis grouping. We also look at co-morbidities and the severity of the illnesses based on the HPD and the risk stratification that we have done and the impact of product. Certainly, geographic location also has an influence. There are a number of other variables that might go into a particular measure that we are calculating. For example, for inpatient services we look at whether the facility is a teaching hospital.

How do we select the measures that are used? Certainly one of the big drivers is the data that are available. There are a lot of things that we would like to be able to measure that we are not able to because we do not have the appropriate data. We also receive input from our internal quality improvement committees. Our medical directors and outside peer advisory committees give us information or suggestions about what we ought to be measuring. We have panels of outside physicians in each specialty who advise us on various matters such as the kinds of data that are appropriate for us to measure each of the specialties against.

Some of the challenges to the collection of data at Aetna U.S. Healthcare are the number of different products and their associated administrative systems. In our HMO, approximately 3 years ago Aetna had upwards of 16 or 18 different claim systems in use. Consolidating that information into one usable format has been a big challenge. Certainly, different coverages bring various kinds of information into the equation.

One of the things that we have to deal with is the completeness of the available information. As I mentioned, we have a number of capitated arrangements with PCPs but also with provider organizations, and in order for us to do the kind of performance measurement that we would like to perform we need to get some feedback from them about what they do. One of the areas that we are very interested in and need their support on is with encounter information. In terms of the volume of data, in order to do valid performance measurement, getting down to an individual provider level, you obviously need a sufficient volume of business in order to make valid performance measurements. Some of the other things we all have to work with are the cost of the data, obtaining the data, and being able to put it into a usable format.

We have an extensive process that we go through to ensure that we are working with complete and accurate information and source systems. We go through a transformation process that loads data into our warehouse. There are edit checks that spin off information about errors or apparent errors and other kinds of anomalies that might be present in the data. Those things have to be checked and validated and corrections made.

How are the results used? Certainly one of the major uses of the information is for provider education. We spend a lot of effort trying to communicate the results to our PCPs, pharmacies, and specialists to provide them with individual results that show them how they compare on various measures against an expected average. Joel mentioned that one of the excuses that providers might use as to why they would not pay attention to the profiling data are that they have a sicker population. Those things are taken into account because of the way we case-mix-adjust the data. We send out report cards. We probably sent out more than 100,000 report cards last year to PCPs. We also send report cards out to specialists and pharmacies to let them know how they are performing.

Another area where we use these results is in our PCP compensation. One of the elements of the compensation model that we have is to pay an extra distribution based on the quality of care delivered. The results that we come up with as a result of these measurements are scored for each provider and become one of the measures in determining what we pay to PCPs. That same idea could also be applied to pharmacies and specialists. The information is also used in our disease management programs. We have extensive programs established for asthma, diabetes, cardiac care, and low-back pain. This information is also one of the elements that is used by the re-credentialing committee.

I mentioned report cards. We are looking at immunization rates for flu and pneumococcal. We are looking for things in the cardiac area, such as the use of lipid-lowering drugs or ACE inhibitors that might be indicators of the quality of care. We look at what their compliance rate is for those members who are on the drugs. What percentage of the time are they using those drugs? We also have outcome measures. We are looking at admission rates and emergency room rates and other factors that are appropriate for that particular specialty. Table 2 is an example of a drill-down on the process measures for cardiac care. You can see some of the immunization rates and members with congestive heart failure taking an ACE inhibitor. There are, for this particular office, 53% of patients using an ACE inhibitor as compared with 46% for the adjusted average for Aetna U.S. Healthcare.

The results are also used for member education. Members can call up our member services area and obtain information about how our PCPs have scored, especially on the member satisfaction survey, so that they can make a more informed choice about the PCP they are thinking about selecting. We also use these results for provider network management. The more we know about our providers, the better decisions we can make in terms of our network.

TABLE 2
CARDIAC PERFORMANCE REPORT

Process Measures	Office	AUSHC
10. Influenza vaccination rates in members 65+ with cardiac disease	60.4%	64.9%
11. Pneumococcal vaccination rates in members 65+ with cardiac disease	36.4	53.9
12. Annual lipid screening rate for member with cardiac disease	N/A	34.5
13. Use of lipid-lowering drugs for secondary prevention of ischemic heart disease	28.4	23.1
14. Members with congestive heart failure taking an ACE inhibitor	53.1	46.3
18. Compliance rate (percentage of time with medication) for members with congestive heart failure	81.4	78.7
19. Coumadin use in members with atrial fibrillation	77.8	42.0

What are the results? Performance measurement is not the only thing that can be done to influence behavior and produce results. Disease management programs are also important elements. Chart 2 shows some of the results of the congestive heart failure patient-care-management program. This analysis includes the results of the impact of a number of different programs that are intended to try to influence cardiac care. In the upper left-hand corner you can see a reduction in acute-care days per 1,000 of almost 20%. For this particular study we were observing high-risk patients who had cardiac care or congestive heart failure. Use of ACE inhibitors is a good sign. We can see that before we started putting programs in place, the use of ACE inhibitors was about 52%. It climbed to 69%. The medication compliance rate, which measures how often patients were using appropriate medications, climbed from 75% to 82%. From time to time we have conducted surveys of our members based on quality of life. We have seen an improvement in those scores too.

At least one of the conclusions that we have come to at Aetna U.S. Healthcare is that performance information can help drive us, as well as providers and employer groups, to look for opportunities to improve care that is provided to our member population. We want to identify members who are most in need of specialized care so we can direct our efforts in that regard. We can determine the outcomes and cost versus benefit of various programs that we put into place and do our best to influence those results.

Mr. Hoffman: Dave, I have a question. How long have you been doing this at Aetna U.S. Healthcare? In general, how has it been received? And lastly, on a broader basis, can you quantify the impact in terms of quality? You showed an example with congestive heart failure that the good performance indicators have

gone up, but in total report-carding over time and in terms of cost-effectiveness have you seen any impact of the profiling?

Mr. Josephson: USQA has been at this for about nine years. As you can imagine, it has been an evolutionary process. We focused most of our efforts for the first five or six years on PCPs. As we get more information, we improve in terms of the measurements that we are performing. We now measure specialists. Pharmacy measurement is one of the more recent profiles that we have added, as well as the inpatient hospital measurement system. We have seen, for those programs that have been in place the longest in terms of sustained effort and outreach, some pretty good results. Obviously, for some programs that have been introduced just recently or have not been focused on, it is too early to tell what those results are going to be. But, obviously, I think everyone here wants to improve quality of care. We think that improved quality of care can help control medical costs, so that is why Aetna has made a big effort and investment in this particular arena.

Mr. Thomas D. Snook: I enjoyed the presentation, and I have a couple questions. It seems that with physician profiling, getting physician buy-ins is a key component. If they do not accept the results, they will ignore what you are telling them. Your technique was interesting because you have a severity definition within the diagnoses groups. I was wondering if you could elaborate on how you developed and defined those severity definitions and how you went about developing your case weights that took those into account.

Mr. Josephson: In terms of how we came up with severity of illness, Aetna has quite a bit of data, so we used a number of years' worth of that data to develop the measures. We used logistic regression models to look at and segment our population into essentially two groups. We developed our model using all the information that we have available that we thought might have any impact on or be an indicator of severity. We input all that information into the model, run it through, develop our coefficients, and then test the results against the other part of the population to see whether or not they seem to be good indicators. For the most part, we found that it is an iterative process. We do go back and make some adjustments so that we come up with a good model for predicting severity.

Mr. Snook: Do the indicators tend to be clinical or were they claim-dollar driven?

Mr. Josephson: They could be either. It could be utilization. It could be drugs. In terms of identifying members and the severity, we rely on different weights with respect to, for example, pharmacy. Maybe pharmacy would be a better predictor if a member is taking a particular kind of drug. Maybe claims would be a better indicator. Maybe some combination would be appropriate. We are

looking at adding lab results into our model. There is a clinical emphasis in our model.

Mr. Snook: Good. And my other question is that I noticed you were using this for PCP compensation, but not using it for specialist compensation. What was your reasoning behind that? Is that something you are looking at down the road?

Mr. Josephson: It is under active discussion within the company and has been for some time. I guess we are just not ready to launch it.

Mr. Martin E. Staehlin: You talked about report cards that you send out and use for provider network management. Does anybody ever get an F? And, if they do, what do you do about it?

Mr. Josephson: We do score providers, and some score very well. Some score relatively poorly. Our emphasis is more on trying to educate the providers and improving the quality of the care they are delivering. In the instance where we have poor performance identified, our medical directors will make a visit, see the PCPs, and review the information that has been developed. We will give them details about the members who we think were not handled as well as they could have been to try to set up a frank discussion with the providers.

Mr. Timothy M. Ross: I am curious about profiling individual physicians versus groups with regard to credibility issues. I think with reasonable panel sizes, maybe 300-500 members, you can probably get some aggregate indicators. Some of your indicators were showing 6-7% prevalence rates for congestive heart failure, which with a 300-500 member panel is only going to be a relatively small handful of patients. If there is an adverse event, it strikes me that it would be hard to distinguish. Bad outcomes sometimes just happen versus a situation where a better quality of care might have prevented that outcome. I am wondering what sort of comments you have in terms of credibility in individual versus group performance.

Mr. Josephson: Certainly one of the things you need to be sensitive to, and we certainly are, is the credibility of the information. For some of these measures you certainly do need to have a lot of information. I think what we found is that, at least in our networks, most of the care is concentrated in 10–20% of the providers. That is really the most credible area that we can focus on. There are a number of providers for whom we do not have enough information to rate them credibly. We are very sensitive to the issue that you raise, and we do look for minimum volumes in order to be able to appropriately perform some of these measurements.

Mr. David L. Terry, Jr.: I am going to take a different approach. Most of the time when we come to these meetings we get more into the actuarial side of the discussion and talk about how we turn data into information. One of our strengths is being able to prepare information, cost-adjust it, and present it in a form that is useful not only from a management perspective but also for use with the providers that are out in the field. For the last four years I have been working for a physician organization and a practice management company, so my clientele now is 100% physicians and hospitals, and the majority of time it is physicians.

We use information and data to help achieve the goal of profiling, which is twofold. We have mentioned it several times now but I would boil it down into two pieces. One goal is to improve the quality of medicine that is being delivered within your network system. The second goal is to stay within the pricing and cost constraints that we are all forced to live with and be able to generate an internal profit and, specifically in my case, generate a profit for the physicians.

There are a lot of dynamics that go on in the industry. When you are in a single health plan it is hard to focus on the dynamics at the provider level. Some of my comments today are coming from that perspective. How does profiling impact contracting and network development issues? How do providers react when they get reports in their hands? One of the obvious issues is that most providers, especially providers in large cities, do not have just one contract with one payer. They have multiple contracts; therefore, they are receiving multiple reports on a monthly basis. It never really dawned on me until I went into an office and started looking at all of the different profiles that come through. From one health plan they get an A. On another health plan they get an F. All of us think we do a very good job at turning data into information, but when you sit in an office with a physician across the table and he or she says, here are 4, and in some cases in larger cities, here are 15 profiles, that physician thinks he or she is doing a pretty good job of practicing consistent medicine. You tell me why 15 different reports give me 15 different answers. That is the type of dynamic that we have to address as people who turn data into information.

We do not have access to all possible information. All we can do is work the best we can. I am going to focus on the keys to successful provider profiling. It is not just taking a report and delivering it to a physician, hospital, or provider and expecting them to just jump out of their seat and say, "Wow! This is wonderful. I am going to change my behavior. This will make everything wonderful and we all will make a lot of money." There are a lot of extraneous circumstances that will impact the profiling system itself. Some of those items I am going to talk about in more detail.

One item is aligned incentives. Without aligned incentives, I do not care how strong a profiling system you have or how well it measures. If you do not get the physician, hospital, or provider buy-in to your profiling system, it does not make any difference how good or accurate it is. The provider has to have trust in you and in the profiling system that you are providing to them. Developing that trust and buy-in is not just giving them a profile report that says how they are doing. We operate in many states and markets. We have 180–190 individual practice associations (IPAs) that we break down into smaller individual accounting units. It is amazing how we use the same profiling system. In one area we do very well with it, and in another area we do not do well with it. It boils down to that trust and the faith that the providers have in you as an organization to work with them to help them achieve their goal, which is basically financial security.

Let's discuss the topic of aligned incentives. The reason I want to talk about physician organized delivery systems (PODS) is because they build credibility into the results you are providing with the profile reports. To give you a simple example, assume you are in a large city and have one large financial reporting unit or one large individual physicians association group that you are working with. You have clumped all the physicians into this huge group where all the physicians live or die with one large risk pool. Take a city I am familiar with, like Dallas, where one side of the city is 60 miles away from the other side of the city. If you have groups of physicians on one side of the city who are being impacted by the practice patterns of physicians on the other side of the city, when you give this physician his or her report and it says how he or she is doing, that physician is likely to say, "I am doing very well, yet there is not any surplus left over in the pool." The credibility of the fact that you are telling the physician that he or she is doing well, yet there is not any surplus generated, does not mean anything to him or her.

What we have found is that we have to break the profile down into financial and reporting units. We have to develop our profiling system so that there is an accountability, specifically since almost all of our models are on a gatekeeper basis. We seek to develop small units of accountability in which the incentive for the PCP is to know all of his or her membership that is being referred to specialists. That is what he or she is accountable for. We set up mechanisms to facilitate risk-sharing and set up incentive pools or contingency pools to make sure that all of the various components stay healthy. It is akin to pooling across a number of different financial entities to figure out how to keep them all in balance.

Subcapitation is something that we also believe in for not only PCPs but our large single specialty panels where you have high utilization and cost patterns. You can identify those specialties that produce 70–80% of your overall healthcare expenditures and get them on a capitated basis. That is a change in behavior that is reinforced with the profiling system. It shows them the results of

their capitated business and the results of other business that is not capitated so they can see the impact of those services.

When we set up our surplus distribution or risk-sharing at the end of the year it is not all on a financially oriented basis. We make sure that a certain percentage of the surplus is distributed based on the criteria that Dave Josephson was talking about. The criteria are based on quality measures, but we also make sure that one of the measures gives us the utilization information from their capitated members. If they are not providing that to us to enhance our ability to do reporting, then their surplus distribution is proportionally reduced. There are incentives for them to make sure that they are submitting data and that the data that are coming back to us, as a management group as well as to the providers themselves, are as accurate as possible. This leads to the surplus distribution and making sure that when those surpluses are distributed they are distributed to the appropriate players. They go back to all the physicians who generated better quality higher member satisfaction and positive risk pool balances. We have found that it is important to give key physicians incentives.

We found that the best way to deliver profiles is not through the mail. It is not through some operation finance manager who has moved into managed care and does contracting through provider relations. It is through a physician-to-physician basis. We try to deliver the primary care profile every month based upon the previous month's activity. The specialists get a report at least once a quarter. Internally, we look at information that is a management tool that we look at every month to find outliers. We can do something immediately if we need to. We do not wait until the providers get their reports. We go out and talk to them. It makes a lot of sense that there are internal reports that are a lot more sophisticated for those who need to be able to analyze what is really going on in the health plan versus those that you deliver to a physician. If you go out with too complicated of a profiling to a physician or a hospital, although hospitals are a little better because they hire people who have more analytical skills, they have a hard time with all the key components. If it gets too complicated they just put it aside. What works is a physician-to-physician interaction.

We usually recruit one of the physicians right out of the group. We make them a part-time medical director who works with our full-time medical director, but instead of having an outside person who walks in with these reports and says, "You are an outlier," we actually have a peer of that physician who is talking to somebody he or she is working with on a day-to-day basis. That peer is in practice and is actually sharing in those risk pools. He has a vested interest. Those physicians all know that that physician who is taking time away from his practice is getting a higher percentage of the surplus distribution as a key physician. But they also know that it is one of their peers. It is not somebody who does not have a vested interest in the organization.

We found that the single most important criteria for achieving change in provider behavior, both from a quality perspective and achieving the financial results necessary or the utilization changes, is the ability to break into smaller groups where we have physicians dealing with physicians and delivering the profiling reports. When the physician thinks that you did not risk-adjust properly or that their case mix is worse than average, they just crawl under the carpet and say, "You are right." Once we get through with this there still is a problem that we need to deal with. I have a couple examples we will go through at the end of my presentation to show you how some of these reports have been used to change some of those behaviors.

The key is frequency. Profiling cannot be done once a year or even once a quarter except for some of your specialists. If you are dealing with PCPs, profiling has to be a monthly activity. For anybody who falls on a scale that looks like they are having some behavior outside of acceptability, we have to get somebody there to talk to them. The profiling reports need to be accurate. Accurate does not mean just for one month. Reports need to be consistently accurate, especially when you are providing them on a regular basis. It is extremely important that what you deliver this month is not dramatically different from last month. If you make a mistake once your credibility is gone for a long, long time. Remember that physician who is dealing with ten health plans and has ten profiling systems on his or her desk? That physician ranges from poor to excellent. Which one is he or she going to believe? He or she is going to believe the person who is his or her peer—the one who is in the physician's office.

I can relate it to the pharmacy industry. The pharmacy industry is probably most successful because they are in your face. There was a presentation yesterday that said that there are 56,000 drug representatives visiting physicians. They are there on a daily or weekly basis. The physicians know who these people are and learn to trust them. When you are delivering a profiling report you are just one of many reports. But to build the trust of the physician it takes a real concentrated effort to make sure that the data and the information you are using gets translated into something that they believe and can trust. That is the reason I truly believe that what we do internally has to be much more sophisticated than what we deliver externally, but we have to be prepared to answer those detailed questions.

I have a couple of examples so that you can get an idea of the impact that profiles have made. Look at physicians 6, 7, and 8 in Table 3. I have left out a few columns and combined some, but the data are adjusted for reinsurance. This table shows the fund status balance—whether a physician is in a positive or a negative position. There are many reports that support this analysis. This is an IPA that we had. The medical director went and talked to physician 6, who

was capitated, and physician 7. Physician 6 said, “I know how to use this system. The patient comes in. I don’t do any work. I get my capitation payment and I send them down the street.” That is the reason so many patients were going to a specialist. The specialists’ costs were way out of line.

TABLE 3
 SAMPLE IPA, INC.
 FUND STATUS REPORT—MEDICARE
 JANUARY 1, 1994 THROUGH SEPTEMBER 30, 1994

PCP ID	YTD MEHB MTHS	Medical Service Fund	YTD Total PCP Cap Paid	YTD Spec/ANC Cap Paid	FFS Budget	FFS Paid Claims	YTD IBNR Claims	Fund Status	Fund Status PMPM
6	160	20,446	3,680	13,498	3,288	11,091	11,544	(19,367)	(121.04)
7	40	4,565	920	2,950	696	9,759	10,190	(19,284)	(482.10)
8	160	20,186	3,680	11,758	4,748	9,345	9,749	(14,362)	(89.80)
9	509	65,306	11,661	7,917	45,728	17,891	18,821	9,218	18.11
11	212	27,341	4,876	1,791	20,676	581	613	19,473	91.88
14	29	4,427	667	498	3,238	414	461	2,418	83.38
21	75	9,747	1,725	1,475	6,547	47	49	6,451	86.01
Total	2,405	311,937	55,292	82,328	174,317	75,926	79,136	19,148	7.96

Physician 8 was the partner of our medical director and was coerced into the provider panel. He did not want to participate. He got outvoted by the physicians to join the IPA. The medical director went in and showed him a few reports, and the physician knew that his profile was out of line. We were not telling him anything he did not already know. He wanted to be out of line. The question was, what are we going to do with him? It took a lot of effort.

Table 4 shows the next time period. What you will find is that from discussion and education in a peer-to-peer relationship we started seeing some real changes in the two problem physicians. An interesting dynamic is physician 9, who happens to be our medical director. When he got the report he wondered what had happened. The first thing he assumed was that it was a change in case mix. He thought that his case mix must be worse than average. We peer-reviewed him and found that the majority of the difference in results was because he spent all of his time with the other doctors. He had lost track of the details of his practice. We found that about one-third of the change was due to case mix. About two-thirds of the change was because he just was not paying attention to detail. Table 5 shows that within a very short period of time the behavior patterns changed. The change was achieved using some fairly simple reports, but the majority of the impact was because we had a physician-to-physician relationship.

TABLE 4
 SAMPLE IPA, INC.
 FUND STATUS REPORT—MEDICARE
 JANUARY 1, 1995 THROUGH AUGUST 31, 1995

PCP ID	YTD MEMB MTHS	Medical Service Fund	YTD Total PCP Cap Paid	YTD Spec/ANC Cap Paid	FFS Budget	YTD FFS Paid Claims	YTD IBNR Claims	Fund Status	Fund Status PMPM
6	297	36,089	6,831	13,522	15,736	12,155	3,997	(417)	(1.40)
7	102	11,724	2,346	2,446	6,932	10,810	3,556	(7,433)	(72.88)
8	219	27,785	5,170	8,631	13,984	9,234	3,037	1,713	7.82
9	542	66,636	12,804	29,561	24,271	45,613	15,003	(36,345)	(67.06)
11	321	38,280	7,595	2,607	28,078	2,648	871	24,559	76.51
14	32	4,857	736	666	3,455	1,516	499	1,440	45.01
21	66	7,644	1,518	1,191	4,935	143	47	4,745	71.89
Total	3,311	406,940	77,340	111,840	217,760	130,057	42,779	44,925	13.47

TABLE 5
 SAMPLE IPA, INC.
 FUND STATUS REPORT—MEDICARE
 OCTOBER 1, 1995 THROUGH MARCH 31, 1996

PCP ID	YTD MEMB MTHS	Medical Service Fund	YTD Total PCP Cap Paid	YTD Spec/ANC Cap Paid	FFS Budget	YTD FFS Paid Claims	YTD IBNR Claims	Fund Status	Fund Status PMPM
6	242	32,590	5,566	9,241	17,783	13,381	1,521	2,882	11.91
7	98	11,611	2,254	3,657	5,700	3,672	417	1,875	19.13
8	200	26,695	4,800	7,649	14,246	5,550	631	8,065	40.33
9	277	36,801	6,459	11,729	18,613	16,175	1,838	2,339	8.44
11	252	31,973	6,010	1,728	24,235	8,229	935	15,071	59.81
14	34	5,031	782	405	3,844	700	80	3,065	90.14
21	64	9,038	1,472	70	7,497	12	1	7,484	116.93
Total	2,826	377,016	66,089	98,522	212,405	139,869	15,760	53,722	19.01

Table 6 looks at the specialty practice of gynecology. Focus on physician 6. You can see his collection rate on his capitation versus what he is actually getting paid. Even with bonuses, it is substantially less than everybody else. The key information is his visits, or referrals per member. His are much higher than everybody else's are. His first comment to the medical director was, "Nobody does workups before they refer the patients to me." When he got a patient he did not feel that they were adequately worked-up. After some analysis we found that this was actually the case. The PCPs were not doing the appropriate workups. When he got the patients he felt he had to do an additional visit and lab work. The issue for us was how to get the PCPs together with the specialists and work

out the referral patterns. What needs to be done up-front? This emphasizes to me how important it is to get this physician-to-physician relationship because if all you are doing is mailing out these reports, then this physician knows what the problem is and feels that it is not his or her problem. He or she feels the problem is because people are dumping patients on him. If you just mail the profile out to him or her that means you have to assume that the physician is going to take the responsibility to go talk to all those PCPs and set them straight. That is not going to happen. Unless you get somebody external together with somebody in the group and bring those two people together in a room and lock the door until the problem is settled, it is not going to get resolved.

TABLE 6
 SUMMARY OF CAPITATED SPECIALISTS
 COMMERCIAL PLANS—GYNECOLOGY
 JANUARY 1, 1995 THROUGH DECEMBER 21, 1995

MD	MEMB Per 1,000	Visits/ MEMB	Procs/ Visit	Net \$/MBR	Coll Rate		Coll Rate With 96 Bonus	
					Billed	Net	Billed	Net
1	9.65	1.74	1.23	\$363	79.54%	115.39%	86.86%	126.01%
2	31.03	1.52	1.25	217	115.09	176.49	132.97	203.91
3	47.52	1.46	1.22	241	85.27	124.21	100.52	146.43
4	65.63	1.69	1.59	310	52.88	90.15	59.58	101.58
5	49.72	1.56	2.20	338	51.67	89.61	71.24	123.86
6	52.69	3.67	1.58	453	28.14	54.89	28.14	54.89
7	60.32	2.06	1.94	448	36.92	61.04	46.36	76.64
Total	38.48	1.88	1.42	330	54.48	89.88	62.98	108.91

As you will see, the problem was not only the fact that this specialist was not getting appropriate workups. Procedures per 1,000 were high for Echo exams. Upon further discussion with the physician we learned that he was one of the few in the area who owned his own Echo machine. He thought it was a great service for his patients. He knew he was capitated and he figured that it did not cost the group any more because he just liked doing the Echos for his patients as a value-added service. But it obviously impacts his collection rate when we review his utilization. He did not really expect to get paid for it. No one could understand why he performed the division of fallopian tubes so frequently. One possibility is that he is known for doing the procedure, so he gets all the referrals. That is a situation that is outside of his control because it relates to benefit design. It is difficult for the physician not to perform a procedure when people say that they want to do it.

This specialist had a significant improvement in his results. The referral rate was way down. The discussions we were having with him were paying off. He did

cut back on his sonograms. In fact, he came down within the norm. Still, his rate of fallopian tube division was still extremely high.

In conclusion, profiling is a tool to achieve your goal; that is, quality and financial stability. In order to achieve your goals you need to build personal relationships and trust and deliver quality.

Mr. Hoffman: Dave, is your job any easier being on the same side of the fence as providers?

Mr. Terry: No, it is not because, being an actuary, I am still the hated person. Most of these issues are handled by our medical staff, but I deal with incurred-but-not-reported (IBNR) claims more than these issues, yet it is still hard for me to walk in and build trust. I have to take a medical director with me to achieve any real changes.

Mr. Hoffman: Is there anything that you do based on your experience that payers miss, either a particular measure or a technique?

Mr. Terry: The hard part for providers is that they get profiles from organizations that are sophisticated, and they get profiles from organizations that are rudimentary. It is a dilemma for the providers to try to make heads or tails out of all the profiles. The key to a health plan being successful at profiling is being able to build trust.

From the Floor: I am curious about how you are able to give reports to the physicians monthly. A lot of the measures you have used are claim-driven so how do you deal with the claim lag? The reports that you demonstrated on the screen used 6- to 12-month intervals. If you are updating each month to a period 6- to 12-months long, how much does it really change month to month?

Mr. Terry: Those examples were from an IPA that just started operations. Most of our reports look at the most recent year's development and the most current month. We look at it on a rolling 12-month basis. We have various cuts at it to try to determine if changes are happening because if you are always looking at a rolling 12-month interval you can miss current trends. We try to cut periods and cut them at different times, but you are right—for these examples they are not always looked at that way. They are looked at more on a quarterly basis, but we do get our reports out monthly for our PCPs. We do try to calculate the reserve on every financial group separately every month. The IBNR gets allocated back into these results so that there is a calculation to show the full estimated financial impact.

Mr. Kelly Meyers: As Joel mentioned, we are in the business of disease and outcomes profiling as well as predictive modeling. We apply artificial intelligence techniques and other nonlinear modeling processes to accomplish those goals. We always make sure that the data are correct. The proper content is provided in the proper context.

Chart 3 looks at per member per month (PMPM) charges, broken out by inpatient, outpatient, physician office, pharmacy, and home health. We are looking at 7 clinicians and an average across about 300 clinicians in the overall group. This is a population of about 5,000 diabetes patients. Each of these providers have at least 100 diabetics in their panel. Let me start out with a couple of questions.

What are your thoughts on physician 46 versus physician 8078? Let me make two statements. First of all, this data are severity-adjusted. Second of all, this data are not severity-adjusted. What is your thought? I am being facetious in a room full of experts, but think about how many times we have heard that the data are risk-adjusted or that the data are severity-adjusted and accepted that and gone forward from there. It is easy to accept that when you are physician 46. It is not very easy to accept that when you are physician 8078.

What I would like to do is show a couple of things looking at cost outcomes. Chart 4 looks at the overall average versus this clinician for inpatient charges, outpatient charges, physician office, pharmacy, and home health. As you can see, every category is at least double the overall average. Again, this data are not yet severity-adjusted. The important issue regarding this chart is what I am not showing you; that is, where this clinician stands on process measures. This clinician has double the screening rates, double the foot exam rates, and double the use of hemoglobin A1C tests versus the overall population. From a National Committee for Quality Assurance perspective, a Health Employer Data Information Set perspective, or a process measure perspective he or she is doing pretty well. Chart 5 brings in severity adjustment for this particular clinician. What did not come through on this chart is that physician 8078's average severity is higher than the average but within 2 standard deviations. Now what call would you make on this particular clinician? You know that he is doing well on process measures. His average severity is higher, but his costs are dramatically higher.

Where would you take that analysis? Chart 6 looks at a five-level severity analysis. I will talk about our methodology of how to adjust for severity and look at the overall breakout as far as total cost PMPM by severity level for this clinician. When you are looking at the data and an average severity versus an expected severity, you really need to look at getting into the standpoint of making sure you are comparing apples to apples. Unless you have the population adjusted for an expected severity, the only way to really do a fair comparison is

to look at it by severity level for this particular clinician. Chart 7 is the same as Chart 6, but severity level 2 is broken out and you are looking at PMPM charges for this clinician versus the average. Interestingly enough, for severity level 2 there were still about 15 or 20 different patients in this particular clinician's panel.

The question that all of you are probably asking is, over what period of time was this analysis performed? This is one year's worth of data for this particular individual's panel. It is usually not my deliberate goal when I start talking to a group to cause more confusion and cloudiness, but that was my deliberate goal here. Usually it is a natural side effect of when I talk to people. The take-away message from this series of examples is that it is important to look at how severity is being calculated and to make sure that the data are severity-adjusted.

As I mentioned, we apply a number of different nonlinear modeling techniques, and there are certain advantages of those techniques for healthcare data. Whether you use neural networks, genetic algorithms, techniques like crystal component analysis, or even correlation matrices, the strength of how we apply these technologies is that you do not have to normalize the data. With healthcare data the richness is in the outliers. Our nonlinear processes do not require you to throw out the outliers. You can accept the fact that you have missing data. If you have a population of 100,000 individuals and self-reported data on a health-risk assessment on 5,000 of those individuals; through traditional techniques you are probably forced to have 2 models. One would contain the health-risk assessment data and the other would not. With our nonlinear process we can actually develop algorithms that apply to situations when data are available and do not apply when the data are not available.

What you get is a phenomenon called graceful degradation, in that obviously the more data you have, the more accurate the model. But you can lose individual components of the data and still maintain a good model. There are some critical data elements. All of the data that we work with are data that are available. Our fundamental models are built on data that are available electronically, such as claims information. When we have health-risk assessment information or lab results information, as Dave mentioned earlier, we treat that as found money and add those things to the model to make it richer, but they are not a principal component of the model.

I mentioned that a lot of what we do is time spent up-front making sure that the data are correct in the first place. Our first application of the artificial intelligence technology is to make sure that the data we are working with is right. An example would be doctor 8078. That was real data that I showed you. It was not just all his data. The data that was attributed to that doctor originally was not all of his patients. There were 100 patients in there, 25 of whom were not his but were actually physician 46's. As you can imagine, the outcomes change

significantly. Everybody talks about taking data quality steps up-front to make sure that the data are accurate. The question I ask is, how rigorous are those steps? We have more than 1,000 different edits that have been generated through nonlinear processes to check the validity of the data. A real live example of that was a hospital organization that was on a 100% peer review because of medical complications that were occurring in the hospital. As you know, 100% peer review is very expensive. When we did our data cleanup process, we identified a glitch in their information systems that allowed a reasonable percentage of patients to be accidentally discharged by the system. They had to readmit those patients electronically into the system. When you do that, you have to have a diagnosis. The diagnosis they were arbitrarily picking was "medical complication." This organization was 100% peer review, not because of the care they were delivering but because patients who were not really discharged were being readmitted. You have to take a significant amount of care up-front to make sure that the data that you are working with is accurate.

Here are some examples of things that we can find. I mentioned that there are more than 1,000 different edits. Thank goodness for us and our clients that there is usually a financial benefit to finding these problems both from the payer and provider side. From the payer perspective, a major concern is claims that were paid that should not have been paid. From the provider perspective, data that was not accurately captured about the particular stay or physician visit is a concern. The bottom line is that about 70% of the time when we look at records we find something of significance wrong just from a financial perspective. This does not even consider the demographic or clinical perspectives.

One other thing about the severity adjustment. On our severity adjustment we look at a dependent variable that is unique for each disease. For example, for myocardial infarction (MI) we look at a dependent variable. In other words, what you are trying to predict is a combination of mortality. We also look at resource utilization and combine those things into a single dependent variable. The reason for doing that is that we are able to come up with higher R^2 s. When you add what we do from a data cleanup perspective to what we do from a choice of the dependent variable, we are actually able to get about double the R^2 s than the traditional severity adjustment models. For something like MI, R^2 s would be in the 0.8 to 0.82 range. We have gone as high as 0.9. Traditional severity adjustment tools get in the 0.3 to 0.4 range, or maybe 0.45. It is a combination of the pains we take up-front to make sure the data are accurate and to pick a dependent variable that makes the most sense for a disease because it does not always make sense to predict mortality for a disease. Is that an appropriate normal outcome or a typical outcome for pneumonia versus in the MI? Mortality should be higher in MI than for pneumonia. That is why we think that severity adjustment tools that only pick a single variable such as mortality or length of stay generate R^2 s that are not where they should be. The way we go about

calculating severity is we will actually identify the dependent variable, and then we will let the technology tell us what the top five or ten independent variables are that are predicting that dependent variable. Then we will validate the models through traditional statistics to give you the walk-through results.

I mentioned that we do predictive modeling, which involves using an individual's health history—by health history I mean claims data or whatever data are available electronically—in predicting that individual's costs or utilization next year or if they are going to get diabetes or any number of different diseases. Some of the organizations we are working with are actually combining profiling with a predictive modeling perspective. Table 7 looks at the variable predicted total cost PMPM for a population of 146,000 individuals. We were asked to identify who would be the top 10% of utilizers next year. They wanted to know for two reasons. They wanted to be able to get ahead of the cycle on those individuals. If you look at this group as a whole, this top 10% of individuals, there is about a 70% turnover in who those individuals are from year to year. The bottom line is that you cannot stay that sick for that long. If your disease management efforts are focused on this year's top 10%, chances are you are going to have different names next year. There will be an appearance of success, but I would bet that the costs are going to be the same because it is a different population. What they are doing from a profiling perspective and from a disease management perspective is taking the individuals who are going to be next year's top 10% of outliers and applying disease management techniques. They are also using this tool, combined with the profiling, to help encourage these individuals' choice of specialists. Which specialists did best for these outliers and how can we help triage these individuals to the appropriate specialists going forward?

TABLE 7
PREDICTING FUTURE UTILIZATION

	Not in Top 10% Predicted	Top 10% Predicted	Total Population
Members	131,452	14,608	146,060
TC \$PMPM	\$59.40	\$398.17	\$93.28
Predicted TC \$PMPM	\$70.67	\$345.33	\$98.28
Admits per 1,000	31.9	190.9	47.8
Days per 1,000	104.48	831.94	177.24
ALOS	3.27	4.36	3.71
ED Visits per 1,000	85.6	214.9	98.5
Cancer \$PMPM	\$2.47	\$25.45	\$4.77
CHF \$PMPM	\$0.03	\$1.21	\$0.15
CV \$PMPM	\$2.82	\$52.94	\$7.84
Diabetes \$PMPM	\$0.24	\$3.38	\$0.55
Endo \$PMPM	\$1.23	\$12.36	\$2.34
Ortho \$PMPM	\$7.34	\$44.17	\$11.03
Renal \$PMPM	\$0.08	\$2.06	\$0.28
Resp \$PMPM	\$5.38	\$28.07	\$7.65

Mr. Hoffman: Kelly, how do your clients use the results of your work, and how do they roll out to providers the profiling information as opposed to the predictive modeling?

Mr. Meyers: Our tools are a combination of reports that look at 10–15 variables that are important for each disease. There is also a data-mining component where you can get to any level of detail. Our tool is usually delivered either through the Internet, where individuals can dial up and use the system on line, or on a CD for populations of a reasonable size. Using a laptop, the medical director can sit down real-time with the clinicians and go through a data-mining process. Assume you are presenting to a pulmonologist. The pulmonologist likely will want to be compared only to other pulmonologists. We have dimensions within the database that allow you to toggle instantly to where they are only looking at themselves versus other pulmonologists. And then the pulmonologist's statement may be, "I do not think it is appropriate to look at aspiration." We can immediately dump those individuals treated for aspiration. This tool allows you to toggle information in and out of the process so that you do not have to go back for a second session with the physician.