AcademyHealth Annual Research Meeting 2007

ROUNDTABLE ON ACHIEVING NATIONAL QUALITY MEASUREMENT & REPORTING PROGRESS & PROSPECTS JUNE 5, 2007

Facilitator

- Eric Schneider, M.D., M.Sc., Associate Professor of Medicine and Associate Professor in the Department of Health Policy and Management, Harvard Medical School and Harvard School of Public Health.

Panel Members

- Elliott Fisher, M.D., M.P.H., Professor of Medicine and of Community and Family Medicine, Dartmouth Medical School.
- Helen Burstin, M.D., M.P.H., Senior Vice President for Performance Measures, National Quality Forum.
- David Hopkins, Ph.D., Director of Quality Measurement and Improvement, Pacific Business Group on Health.

Eric Schneider: I'm Eric Schneider from the Harvard School of Public Health in Boston. Like many of you, I began my trip to Orlando in the back of a taxicab. When I explained to the cab driver that I was going to a meeting of health services researchers to discuss quality and cost of medical care, she became very engaged and discussed all of the problems with American health care, including problems with access to primary care, and the lack of information for patients about the quality of doctors and hospitals and the prices they charge.

I took from our conversation that the public is beginning to get the message from the Institute of Medicine's (IOM) *Crossing the Quality Chasm* report. The public is also beginning to perceive some of the solutions regarding performance information and information on the cost of care.

Our session on national strategies for measurement and reporting is a nice conclusion to a meeting that has featured performance measurement and reporting as a fundamental arch in the bridge that will take us across the quality chasm. Almost every presentation during the Annual Research Meeting has referenced performance measurement and reporting as tools that will be employed in this enterprise of quality improvement.

Measuring and reporting on performance has been around since at least 1986, when HCFA started the hospital mortality reports. In the 20 years since then, there's been slow but steady progress in developing measures for hospital care, long-term care, health plans, and now, increasingly, physician groups. While many health plans and private entities are using measures to pay for care, and pay for performance, there are still skeptics about these efforts.

Some anxiety about performance measurement reporting has resurfaced among providers regarding Medicare's Physician Quality Reporting Initiative (PQRI), which expands the number of quality measures and starts paying providers based on performance. The question that intrigued me and prompted this session was whether this performance measurement reporting "arch" we're ready to step out onto is sufficiently solid. Will it bear our weight as we try to move across the bridge and across the quality chasm?

To address this question, we have three distinguished panelists. Elliott Fisher is Professor of Medicine and Community and Family Medicine at Dartmouth. He was also on the IOM Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs. Helen Burstin is Senior Vice President for Performance Measures at the National Quality Forum (NQF). And David Hopkins is Director of Quality Measurement for the Pacific Business Group on Health. He is also an affiliate at the Center for Primary Care and Outcomes Research and the Center for Health Policy at Stanford University.

Elliott Fisher: Let me start with a story. Two weeks ago my mother was hospitalized with chest pain. She was taken to a hospital without a catheterization lab and experienced a 6-hour delay between the onset of chest pain and receiving her coronary angiogram, which did show significant occlusions. Physicians inserted stents and eventually averted a major anterior myocardial infarction. Her ejection fraction is back to normal, but the initial technical quality of care she received was pathetic.

There was no real harm done, other than to my own sense of having failed to improve the health care system yet again.

She experienced four hospital transitions, moving back and forth between an acute care hospital and a rehab hospital. On three out of four

transitions, her providers asked me to help with the medication reconciliation. The intern called me to ask, “What medications is your mother on?” And then called back to say, “I’m sorry, we don’t have that medication in our hospital. Could you go to the other hospital and pick some up for us?” It was a remarkable portrait of a badly broken health care system, especially for patients suffering from chronic illness.

And so it is a good time for me, personally, to look back at the work of the IOM Committee, and for all of us together to look forward, and ask, “Are we getting where we need to go?” Our IOM Committee on Redesigning Health Insurance Performance Measures, Benefits, Payment Programs, and Public Reporting renamed itself, for simplicity, the Committee on Redesigning Health Insurance. The Committee convened at a time when we believed the gap between the current care and the vision of the IOM’s *Crossing the Quality Chasm* report remained wide.

I would like to ask you as we go through these recommendations, which are now 2 years old, how much progress have we made in the last 2 years? We concluded that there is a huge gap between the current system’s high costs and poor quality and the safe, timely, effective, efficient, equitable, and patient-centered care that the *Quality Chasm* report envisioned. We believed, as it seems almost everybody in this room believes, that performance measurement is actually the key to further progress. Performance improvement seems to depend upon good, reliable, robust performance measurement whether it’s pay for performance, consumer engagement, or quality improvement initiatives that would help providers improve what they’re doing.

We concluded at the time that recent public and private sector initiatives on performance measurement were very important but unlikely to achieve the national performance measurement system we need. We articulated concerns about a narrow—and the word I found when I reread the report—*timid* focus of measurement, largely around technical quality. In addition, duplicative and inconsistent measurement systems had proven burdensome to providers, who were expected to abstract data from medical records. We expressed concern about the independence and integrity of data and whether funding levels provided to achieve a robust measurement system were adequate to support the data’s reliability and independence.

First and foremost, we recommended establishing and funding a National Quality Coordination Board (NQCB). We envisioned a separate agency within Health and Human Services, with board members serving staggered terms set far enough apart from the political cycle to ensure the Board’s independence.

The NQCB would serve several purposes, including: (a) setting clear, short and long-term goals for improving the health care system; (b) designating or developing standardized performance measures that could be widely implemented and adopted; and, (c) ensuring the creation of a robust data repository system, capable of reliable, valid, efficient data collection at the patient level.

Patient-level data collection would allow us to examine the performance of physicians, hospitals, and nursing homes where patients were getting their care. Such a data collection system would facilitate community-level analysis so that we could better understand disparities, the care of the uninsured, and care of racial and socioeconomic groups who are vulnerable in those communities. With a robust data repository, the NQCB should develop and release useful reports responsive to all stakeholders, understandable by consumers, and useful for physicians and other providers trying to improve the quality of care.

The NQCB should identify and fund a research agenda. We felt many measures were missing, and substantial investments were going to be required in order to achieve the performance measurement system we needed. The NQCB would evaluate over time how performance improves related to measurement and how to improve policy.

The IOM Committee endorsed a starter set of measures as Congress requested us to do. We endorsed a number of the measures that had been developed by the Ambulatory Quality Alliance (AQA) and others, including consensus-based provider-specific measures for nursing homes, hospitals, or physicians.

We endorsed a new measure calling for 1-year mortality rates for acute myocardial infarction (AMI), along with more mortality and functional outcomes a year after the AMI. We called for the development of an array of new measures that would examine the experience that patients have over time with care transitions, and that would foster shared accountability among providers.

Where are we now? The answer is discouraging. In September 2006, the Committee on Redesigning Health Insurance recommended moving forward with pay for performance but called again for measures of quality, cost, and patient-centered outcomes. The report underscored the need for implementing pay for performance in the context of a “learning system” — where different approaches to performance measurement and payment reform would be rigorously evaluated, allowing us to learn from experience and to improve our approaches—as we move forward.

Congress has not pursued our recommendation for an NQCB. We remain dependent upon a poorly funded and largely voluntary consensus-based

process to achieve the performance measurement system called for by our IOM Committee.

Kevin Weiss and I now co-chair a committee charged with developing a measurement framework that would encompass quality and costs and patients' experiences over time. Our first case example focuses on the care of patients with AMI and emphasizes the importance of following patients over at least the first year after their initial event. Measures would include key processes (including the degree to which care was aligned with patients' needs and preferences), health outcomes at 1 year, and the total costs of care. These measures would provide a framework for evaluating the relative efficiency of different delivery systems and regions. Another pilot area will be back pain.

Where might we think of going? First, we need to think carefully about what will be required to put in place an effective national performance measurement system. There is great public and policy interest right now in comparative effectiveness research. An infrastructure for such research would encompass both randomized trials and observational data. Research needs to improve our understanding of whether the drugs that are being introduced lead to better outcomes and whether we can identify quickly the specific risks of those agents as they're introduced. The notion is to compare treatments, controlling for patients and providers.

The performance measurement system envisioned by the IOM Committee would be similar in that it would compare providers while controlling for patients and treatments. It's essentially a data system that examines outcomes, costs, and how these vary across providers. In 10 years, a well-designed information infrastructure should enable us to compare providers and evaluate whether pay for performance makes a whit of difference.

Helen Burstin: I'm going to give a slightly different perspective, one more NQF-centric, but one that recognizes the current reality, which is an alphabet soup in terms of quality measurement reporting.

I agree with Elliot Fisher that the lack of a national goal-setting process is problematic. We're in this very strange place of not having enough measures in some areas, and having way too many measures in others. There are many overlapping measures that do not harmonize across providers, and many measurement gaps exist. We have too much process and not enough outcomes, and we are working hard to adjust that balance. Methodological issues abound, and I'll discuss those in a moment.

One emerging issue is the challenge of distinguishing between basic competency and high performance. The most recent round of pay-for-performance initiatives by CMS has brought the physician community and

others to the fold in terms of trying to develop measures. The attachment of the measures to financial incentives has, however, at times pushed the bar too low. We need to think about how to move beyond some of those basic competency measures, perhaps in the next round, to push towards higher performance.

So where are we with quality measures? On the positive side, the measurement development and endorsement process has activated health care leadership. People are engaged, and that is probably a positive sign, as we think about moving forward. Harkening back to Elliot Fisher's comments, we critically need a national quality measurement reporting infrastructure to set priorities and goals, to develop and maintain measures, to endorse "best in class," and to address the problem of multiple measures on the same topic. We need this infrastructure to harmonize our disparate efforts and to pick measures that we can all be measured on, instead of the cacophony we have at the moment. We need to link process measures with outcome measures. Measures need to address episodes of care, and the efficiency, cost, and value within those episodes. We continue to have significant measurement gaps, particularly around patient engagement, care coordination, and self-management.

Attribution is another important issue. We need to better understand attribution at the individual physician level. The issue of measure exclusions continues to come up. We are clearly in a transition period to electronic health records (EHRs). At the same time, we are seeing measurement drawn from new CPT II coding, hopefully increasing the chances of robust measurement through administrative data.

I would argue that health information technology is probably one of the ways we could solve a bit of this mess.

There are multiple efforts underway to specify the key EHR data elements needed to produce quality measures. NQF has a small piece of this, specifically understanding the quality of data elements needed to secure robust measurement. With interoperable systems, we can also begin to understand some of those high-impact gap areas, for example, transitions and coordination of care.

We also need to do a better job at measuring disparities. Measures of health care disparities are essentially quality measures, and we need not marginalize them but link them to quality. We should always be looking at our data and thinking about which measures should be routinely examined for disparity. NQF's ambulatory care project is developing a set of principles to identify measures that are particularly disparity sensitive.

I want to mention population health as another area we must not lose sight of as we move forward. As we're building the system of the future, we

need to include population health as a priority issue. We're now seeing an emerging movement around measures of public health quality. We need to consider how we align with that.

David Hopkins: I believe all of us have the same vision, and we offer different perspectives. I am a big fan of the IOM report on performance measurement. The report recognized that there are, in fact, multiple users and multiple uses for performance information. It found that we need performance information for providers—physicians and hospitals—to lead improvement activities; for consumers, who are interested in learning about the quality and efficiency of care; for employers and health plans, for payment and benefit design; and for public health authorities looking at population health issues.

I believe that three out of four of these user groups have been largely neglected in the work that's been done to date, so I would like to offer the consumer and purchaser perspectives on performance measurement.

The Consumer Purchaser Disclosure Project, funded by the Robert Wood Johnson Foundation, is an organized group of large employers, labor unions, Consumers Union, the AARP, and employer coalitions. This effort has been underway for 3–4 years. The Project has concluded that the pace of development and the scope of measures that are on the table currently are too slow and too narrow. In the meantime, costs continue to rise, and patients continue to make poor choices because they don't have the information they need.

We need a robust performance dashboard, because consumers need this information. We certainly need such a dashboard if we're going to do intelligent work on plan design and benefits. And finally, we need information to be valid and rapidly available. Our mantra is "don't let perfection be the enemy of the public good."

How do we assist the AQA, the AMA Physician Consortium, and NQF in identifying what measures are needed? What criteria should be used in judging those measures? We recommend judging measures based on reasonable scientific acceptability, with an emphasis on "reasonable." If we're striving for perfection, we will not get far, because there's not a single measure out there today, even those fully endorsed by NQF, that is absolutely perfect. We have to figure out where to draw the line on what's acceptable and feasible.

And, then, the whole relevance issue is about meeting the dashboard that the IOM created, giving consumers their choice among all kinds of conditions and procedures, and demonstrating the high variation in performance. Does it affect large numbers of patients, or large amounts of money? "Outcomes, outcomes, outcomes" is another mantra of ours.

Arnie Milstein and I recently put together a scorecard on where we stand with nationally endorsed performance measures today. The scorecard views the end stage as a full and useful set of measures for the purposes that I described. We organized it around the IOM's six domains. We broke the effectiveness domain into process—which includes timeliness—and outcomes, with safety, patient-centered, equity, and efficiency remaining as separate components. Then we imagined what would comprise a robust measure set that we should be aiming for.

For process measures, we need to cover as wide a set of conditions as possible, and we need to capture coordination of care. On outcomes, we need mortality, morbidity, and functional health status for a wide set of conditions. For safety, we need both structural measures, like safe practices, but also outcomes. What are the infection rates that are occurring in the course of treatment? The AHRQ Patient Safety Indicators give us a start there, but they're not nationally endorsed right now, at least not by NQF.

In terms of measures of patient-centeredness, credit is due to both AHRQ and NQF for bringing us the Consumer Assessment of Healthcare Providers and Systems (CAHPS) family of patient experience surveys. The CAHPS survey is covering many important domains and has given us a good start in capturing patient centeredness. We're still missing some significant domains, including shared decision making, coordination of care, and measures of equity, safety, and errors.

Our scorecard indicates that we're doing a good job on process measures as far as they go. Many conditions are not addressed by these measures, however. In terms of outcomes and safety measures, our scorecard shows we have a long way to go.

While we're struggling nationally to gain acceptance for consensus-based measures, independent efforts at the local and regional level are moving ahead rapidly. These regional activities should spur us to move more quickly at the national level. For example, the Leapfrog hospital rewards program combines NQF-endorsed quality measures for hospitals with resource use efficiency measures that have not been endorsed yet.

Bridges to Excellence is using systematic office processes and demonstrated excellence in three clinical areas for its rewards program. In Minnesota, HealthPartners is looking at diabetes care from the patients' point of view. HealthPartners recognizes that it does not help patients to see a physician who always remembers to order the A1C test, but never the cholesterol test. The question they are answering is, "How many doctors are actually doing *all* of the important steps in diabetes care?"

In Massachusetts, the General Insurance Commission has compiled data that examine physician-level clinical quality and cost efficiency using best-in-class vendor tools. Some states are undertaking hospital infection reporting and doing a remarkable job of improving their understanding of hospital infection rates and how to reduce them. These efforts are using measures that are not yet nationally endorsed.

I want to highlight for you a great opportunity that is before us, the Better Quality Information Pilot Program sponsored by CMS. Six states are involved in the program. The purpose of the pilot is to aggregate data and bring Medicare data together with commercial data. By aggregating data, we can look at all the care that physicians are providing and, ultimately, do a better job of measuring that care. At the end of 2 years we'll have 12 quality measures collected in two consecutive measurement cycles. The purpose of the program is to demonstrate how the public and the private sectors can work together to combine data and create a large enough sample for measurement.

I see three dynamic tensions in the movement for physician performance management. Providers rightfully need actionable measures. Consumers, on the other hand, want outcomes. So do we create process measures that are actionable, or are we seeking outcomes measures? The answer is clearly both.

The second tension is one that's not widely discussed, but I think it's time to give it visibility. Providers expect high precision. They all went to medical school and studied the scientific approach. They read journals where the standard is 95 percent confidence intervals for new treatments, so it seems natural to apply that standard to classifying providers. Consumers, however, aren't willing to accept the 95 percent confidence interval because it means that there's no differentiation. They prefer to have more differentiation.

The third tension is feasibility, and the limitations of measures that cannot be produced today without incurring the high cost and burden of manual chart abstraction versus those that can be derived from available electronic data.

Eric Schneider: You've heard three different perspectives on the outlook for performance measurement and reporting. I want to open this up for audience discussion.

Greg Pawlson: Greg Pawlson, NCQA. I was privileged to be part of a group that Janet Corrigan, Sean Tunis, Peggy O'Kane, and a number of others convened recently to ponder these issues. We focused on four issues: (1) funding, (2) the need for a framework, (3) fusion of data systems, and (4) research on comparative effectiveness.

First and foremost is funding. We've got what we've got because we don't put any money in this. NQF is woefully under-funded. NCQA is always

holding out the tin cup trying to find money to create measures. Without sufficient funding, we have a measurement system that is woefully lacking.

Second, we need a framework, and NQF is moving towards that. Janet Corrigan has done a great job of bringing her experience at IOM and NCQA together, and creating an organization that is bringing people together around developing a measurement framework.

Third, we need better fusion of data systems. Our current generation of measures reflects the fact that we don't have the data to measure more accurately. One of the things that we've been looking at lately is how accurate the measures are using current data systems. I have a paper in review with the *American Journal of Managed Care*. We looked at HEDIS measures, specifically which ones could be measured at the health plan level within at least 80 percent accuracy.

With a sample of 411 health plans, we determined there are virtually no HEDIS measures with a hybrid specification that can be undertaken accurately at the health plan level, let alone the clinician level. So, it's not going to be 50 percent accuracy, it's going to be probably 5 or 10 percent accuracy with many of the measures that are currently in use. And I think the worst thing we can do is throw garbage at a fan and think we have fruit salad. The fourth area is further research on comparative effectiveness.

My underlying point is that we need to tackle the funding problem or else we'll be here in 10 years having this same symposium.

Helen Burstin: We would prefer not being reactive. In fact, what we really need to do is gain the recognition that this enterprise is a public good or we're always going to be operating on soft money. There is an effort called Friends of NQF that is leading the effort to secure an appropriation of some sort, not just for NQF, but for measurement development more broadly.

Elliot Fisher: A better framework may help us prioritize where to invest our measurement infrastructure. I note the remarkable success with the CAHPS measures, specifically the ability to identify patients cared for by a specific group of providers and collect survey information from those. This framework may help us focus more parsimoniously on a set of measures that do not require a forest of data collection and chart abstraction. That would be very efficient.

Second, as we work together to influence Congress on the funding, I hope that we think about comparative effectiveness research and performance measurement as parallel data collection infrastructures that could support this system.

Earl Steinberg: Earl Steinberg, Resolution Health. We need to stop thinking about competency at the same time we're talking about performance

measurement. For example, a measure that asks did the dermatologist examine the skin? That type of measure should be left to a licensing board or a professional society. My second point is that, in my view, chart review is simply not an option. It is too expensive and it's riddled with methodological difficulties that are as challenging as those with administrative data. So the reality is that for now, given all its limitations, claims data supplemented by lab results is the only game in town.

My last point is that while we'd like to have outcome measures, they're difficult to achieve. In thinking about your mother's experience, Elliot, an outcome measure would have missed what transpired.

Eric Schneider: Only a handful of evaluations have examined claims administrative data compared with other types of data sources that could be used for this purpose. These evaluations came to varying conclusions, and the conclusions were very measure-specific suggesting that research on a broad set of measures is needed.

Helen Burstin: We need to look to the future in terms of health information technology standards and ask, "How do we fix EHRs so that they're ready to incorporate measurement in the next few years?" In the meantime, we've got to figure out how to use administrative data well, and figure out how to use whatever combinations of data processes we can to move forward.

In terms of the competency issue raised by Earl Steinberg, it is important that the physician community is engaged and that physicians report. Hopefully, the measures will begin to correct themselves and improve over time.

David Hopkins: Let's discuss the data quality issue a little bit more. Administrative data are not perfect—far from it currently. Would it be worth spending time and effort to make them better? I think so, as opposed to either waiting for the EHR or introducing a new system that imposes more burdens on physician offices. What if we made the data really count through performance measurement and public reporting, and if we're going to do that, let's not make the mistake of only making them count for under-use measures. We need to include over-use measures, too; otherwise the coding system will be gamed.

There are ways that we could work with the physician community to help them improve coding and documentation systems. At the same time, we can help ramp them up on EHRs as quickly as possible.

Tom McGuire: My name's Tom McGuire, from Harvard Medical School. One of the challenging aspects of this area is that we will want to move forward with quality reporting without having studies on how it's working. I'd

like to make an assertion and then ask a question. The assertion is that all of these quality measures pick up just a fraction of what we want providers to do. So, while there are only a few cancer measures, there are actually a range of treatments hospitals or physicians can offer for management of a chronic illness. This gap raises the possibility or likelihood of unintended consequences. So the empirical question is, if we measure 20 percent but we have 80 percent that we care about that we don't measure, what is the relationship between those things? If we improve in one area that is measured, does that detract from other areas that are not measured? Or does measurement tend to spill over and improve other areas as well? What evidence is there to suggest a positive relationship between things we measure and things we care about? Or do we just cross our fingers?

Eric Schneider: It's a question that's frequently posed by the provider community, and it's interesting to hear it coming from an economist.

Greg Pawlson: Several studies suggest at least reasonable evidence on this question. One was the VA study, which examined incentives given to Veterans Integrated Services Network directors on certain measures. The study found that those measures were performing at a relatively high level, higher than the civilian hospitals in the given region. Where there were no incentives, the hospitals were doing as well as the civilian hospitals.

We have to keep studying, but at least empirically, there didn't seem to be any negative spillover. The second study is from Martin Rowland's work in the United Kingdom which is going to be coming out in the *New England Journal of Medicine*. I think it will be supportive and show no huge effect.

Now, at the other extreme, some may argue that we should not measure at all because of the danger of adverse effects. So I think the point is taken, and it does mean that we need to do more empirical studies.

Elliot Fisher: There are dangers both in moving forward and in not moving forward. This risk underscores the need for funding so that we can both improve and add measures.

Rick Kronick: Rick Kronick at University of California, San Diego. I have a similar question. Everyone is concerned about teaching to the test. There is a reason that we pay apple pickers on the ultimate pay-for-performance system, and that pay for federal judges is disconnected from performance. These payment systems relate to our confidence in measuring accurately, and worries about distortions. As a clinician, and as someone involved in the process of measurement, I wonder what your sense is of that balance, and specifically concerns related to potential distortions and teaching to the test.

Elliot Fisher: The dangers you describe increase as you move down in scale to measuring individual physician performance, and as you move from measuring and reporting to paying for performance. Physicians seem to respond positively and creatively when measurement and profiling is done within the group. Individual reporting at the physician level, linked to pay, risks, serious problems, and distortions.

My anxiety increases as we move toward measurement and reporting at the individual physician level. I argue strongly for measuring group performance and trying to help physicians come together to become reflective practitioners. Moving to outcomes and costs, as the NQF is starting to move, while hard, will facilitate provider discussion about differences in outcomes within practices.

So I advocate moving toward measures that pull groups of providers together to say how we work to achieve goals, and then the measurement within the system can be at the physician's specific level. At least that's the model that might work best in our chaotic system where we don't have good evidence on how things should work, or are working.

Helen Burstin: Since we have AcademyHealth and numerous health services researchers in the room, I would like to add that we have an incredible opportunity to study natural experiments that are happening as we speak. We should work hard to evaluate wherever possible. For example, some of those physician-specific measures got only a time-limited endorsement from NQF, because they haven't been tested.

David Hopkins: Teaching to the test is a very real concern. I was going to bring up another natural experiment, the IHA pay-for-performance program in California. Briefly, this program measures, rewards, and reports at the physician group level in California. We have partnered with NCQA to develop the technical infrastructure for that.

Initially, the IHA program embraced only a small set of measures and it is clear from the evidence that teaching to the test occurred. We're in our fourth or fifth year now and a very interesting thing happened as the measure set became more robust recently. Physician groups realized that it was no longer cost-effective to deal with issues one measure at a time. They started thinking in terms of systems. And, of course, the answer, or a big part of the answer, is to think in terms of systems of care. The experience of the IHA program is one more reason why I would argue for a more robust set of measures.

Ann Elixhauser: My name is Ann Elixhauser, and I'm with AHRQ. I wanted to follow up on something that David Hopkins said, and that Earl Steinberg also mentioned. A couple of years ago, AHRQ funded a study to look

at what additional information could be put on the administrative data set in order to make administrative data more reliable and valid for public reporting of hospital quality. Those studies are just starting to come out. The first study was published in the *Journal of the American Medical Association* this year. The study was conducted by Michael Pine and Associates. Three other studies are on the way.

The conclusion of those studies was that by adding information on whether diagnoses were present on admission or not, and laboratory values on the first day of admission, and potentially vital signs—that adding that information could get you almost as much as a full clinical model as operationalized in that study. The data used in those studies were from Pennsylvania, which relied on medical chart abstraction. Administrative data is probably the only real source of all-payer data that is available across nearly the entire country. What AHRQ has done is to say okay, what incremental improvements can we support right now? AHRQ has released an RFC to support data organizations at the state level, such as state hospital associations, or government organizations. The goal of the RFC is to support those organizations in their efforts to add clinical data to their administrative records, including lab values at admission. Because lab values are available electronically in over 80 percent of hospitals nationwide, we're hoping that this will make some sizable incremental improvement in our ability to measure quality right now.

Jessie Sullivan: Jessie Sullivan, Hudson Health Plan. I wanted to respond to several comments related to competence. Primary care physicians may have as many as 1,500 patients. In terms of current measures for almost all of those patients, we have three or four indicators, and perhaps 20 for patients with a chronic disease. As a result, a physician may have up to 15,000 indicators that he or she is tracking. If physicians can do the basic right thing 15,000 times, they're going to be providing very good care. Giving an eye exam to a diabetic is not a complicated concept. Making it happen in a system where we don't communicate is an achievement.

I wanted to comment on all-or-nothing measures. We have three different kinds of measure sets that we do in our health plan, and for immunizations we have an all-or-nothing measure for 2-year-olds. If 2-year-olds have all the immunizations, then we pay the bonus; if they don't have all their immunizations, we don't pay the bonus. As medical director, the advantage of this approach is that every single immunization is worth \$200. The disadvantage is that we have had instances where a provider has stopped sending reminder notices to patients who may have missed an immunization. Lead screening is another problematic area when it comes to all-or-nothing performance measures. Many of our patients refuse lead testing because it requires a venous draw.

John Horner: I'm John Horner with the Hospital Industry Data Institute in Missouri, part of the Missouri Hospital Association. I would like to comment on administrative data records.

Every administrative record should include height, weight, and smoking status. Coding quality is a major problem, and a lot of administrative data sets are limited by the number of codes they can have. The most I've seen is 61 ICD9 diagnosis codes on a record. The data sets are limited to 10, so these codes are limited in the richness of information that's available.

In addition, we need to follow individual use, where there is a lot of duplication on administrative records. We have patient identifiers on our data, and some patients have up to 126 episodes of care in a year with probably 15 inpatient hospitalizations, and the rest were outpatient hospitalizations for multiple causes, including injuries, drug use, etc. This duplication makes it difficult to examine administrative records in terms of population health.

Michelle Seelig: Michelle Seelig from the Puget Sound Health Alliance, a regional partnership involving employers, physicians, hospitals, patients, health plans, and others working together to improve quality and efficiency while reducing the rate of health care cost increases. We have 15 data suppliers.

Many of the methodological issues I struggle with result because we're using administrative claims data. Claims data simply lack many of the data points we would like. My question to the panelists is, how do we make progress with electronic medical records? We're trying to do this with administrative data, but many of these limitations could be overcome with electronic medical records. The alliance is committed to making this shift.

Eric Schneider: Your comment raises an important issue. CMS is poised to implement a pay-for-performance program based on CPT II code reporting. There is evidence that providers have become engaged in performance measurement, more than they were 10 years ago. If they're burdened with this sort of reporting of claims codes, would their desire to be engaged dwindle? Is it better to design a transitional technology that will lead to EHRs?

Elliot Fisher: I agree. We need to migrate from administrative data to EHR as quickly as possible. Unfortunately, many of us are discovering that the EHRs available on the market do not necessarily solve the problem. You can't just push the button and get the data that we're talking about.

Providers investing in those EHR systems need to let their vendors know that expanding the measurement capability is essential. In addition, a lot depends on the way physicians use EHR, and if they're going to write free text. If they're going to write all the clinical information in free text, that's not going to solve the problem. So a lot of work is going on in this area with Boston

researchers leading the way in terms of studying this problem within the Massachusetts e-Health Collaborative and other projects. Ultimately, we need to merge EHR and billing systems.

Helen Burstin: The last thing we want to do is penalize early adopters. CMS is working to develop a set of specifications for a handful of measures for the PQRI that may be submitted via EHRs.

And some of the vendors are already working with NCQA on physician recognition programs. I think the key will be to include quality measurement reporting in future certification requirements. We're at a tipping point. We will probably accomplish much of this within about 5 years.

Elliott Fisher: A lot of our current measures grew out of our struggle to try to measure what we thought would be some easy things to do. We relied on administrative data initially, or chart review. You have lots of efforts to build into the EHR measures that were developed and specified in this desperate struggle for early measures for pay-for-performance or public reporting at the individual physician level. We risk building into EHRs specifications for measures of performance that were based on research from the mid 1990s and will be out of date by 4 years from now.

I'm nervous about where we are. While we're trying to move forward with measuring the things we know are important, we ought to really step back and think strategically about performance measurement. What constitutes good care? Should we continue with efforts to measure at the individual physician level?

Helen Burstin: We need to consider the new crop of measures that we could potentially get out of a really good next-generation EHR. We're less interested, however, in specifying down to the data element level how to make these measures last forever and be incorporated in EHRs. It is about building the next rocket, not the one we're flying right now.

David Hopkins: The architecture that's needed is already in place. Let's improve the data so that we can both run measures that are currently endorsed and undertake research on the 80 percent of care that's not evidence based, as Elliott Fisher suggested.

Eric Schneider: I want to thank the panelists for an engaging portrait of the current status. I also want to thank the audience for a stimulating set of questions.

Prepared by Kristin Rosengren and Margaret Trinity