

**Pay for Performance: A Critical Examination:
Part II
National Committee for Quality Assurance
December 1, 2006**

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

[START RECORDING]

GREG PAWLSON, M.D., M.P.H.: ...that it's very tough to stop the conversation in breaks. In fact, I said to David Blumenthal earlier, that if at least ten projects don't get launched at this meeting, it will have been a failure. And actually, David and I were already talking about two projects that we're thinking about doing. So I appreciate it, and sometimes the talk between is very important. However, we do want to try to stick fairly close to our agenda.

What we're going to do in the next hour and a half, in two different 45-minute segments, is to bring some data to this. We're talking about basing pay-for-performance on actual data. We'd like to bring some data to bear on the fact of whether or not pay-for-performance is working.

I'm Greg Paulson. I'm the Executive Vice President of NCQA, and I wanted to kick off this segment. And we're going to do it in a little bit different format. We think variety is nice here, and the way we're going to do this particular segment is I'm going to introduce the two individuals, and then they're going to each have a short presentation, and then they're going to engage in, hopefully a very, I think, thoughtful kind of dialogue around the British experience and bringing in some perspectives from our own experience in the

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

United States with pay-for-performance. So, I think you'll enjoy that kind of an interaction.

The two individuals we have to do this, I think, you have their bios. I'd just like to say two quick things. First, in terms of Tim Doran, who's a clinical research fellow in public health at the University of Manchester. I have actually had the privilege of interacting with he and a number of his colleagues, including Martin Roland for the last couple of years, and also revealed to them that I was privileged to review their papers that they've submitted, one of which Tim was the lead on, was published in the New England Journal. And another one that I think is probably eminently going to be published in the New England Journal, on the experience of our British colleagues.

And it's interesting, sometimes you think about Britain as being somewhat cautious. In this case, they passed us in a huge bold leap in the pay-for-performance program that was introduced with the British general practitioners on, essentially a national scale. And Tim's going to describe the results of that, and I don't think we want to call it an experiment, that grand implementation of pay-for-performance. To engage him with that dialogue is Bob Galvin, who is the Director of Global Health Care for General Electric and someone very known to NCQA since he is a board member of NCQA.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

But Bob has also been, I think, a thought leader in health care from his perspective with purchasers, and has really been a seminal person in leading purchasers into thinking in relatively radically new ways about the delivery of health care. And the perhaps prime example of that, that we've worked with is the Bridges to Excellence program, which some of you are aware of, that really was one of the first, and has perhaps had the most far-reaching effects of any pay-for-performance program in the United States. So without further ado, I'm going to turn it over to the two of them and let them have at it.

TIM DORAN, M.D.: Thanks very much. I want to give my thanks to NCQA for inviting me and flying me across. I don't know if what I'm going to be describing is a bold leap or a bold plunge, I'll let you decide. I should just say this program applies to primary care in the whole of the United Kingdom, not just Britain, it includes Northern Ireland as well. And I want to just talk a little bit about some of the background to the program, how the program actually works, and then introduce the results of some of the analyses that we've done on the first couple of years of the program, both the intended and the unintended consequences. And I think Bob's going to pick up and respond and we're going to have a little debate about it.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

First of all, just to say that the whole purpose of pay-for-performance in the United Kingdom, as I expect here, is really to try and improve the quality of care received by patients. But it's important to point out in the United Kingdom, it's not the only attempt that's been made to do this. The Department of Health has introduced several initiatives since 1997 with the same aim. And I don't propose to go through these in any detail, but I've only listed them to indicate that for several years now, physicians have been having their performance monitored and measured and fed back to them, so they were getting quite used to it before the pay-for-performance ever came into being. And there is some evidence that quality of primary care in particular has improved as a direct result of some of these initiatives.

So measurable quality was already on the increase prior to 2003. So in 2003, just to give you some of the background, the Department of Health in the U.K. identified about 1.8 billion pounds, roughly \$3.5 billion, of additional funding that they wanted to commit to primary care. But it was quite keen [misspelled?] that this money was only going to be committed if they could get some tangible results in terms of improved quality care for patients.

So they approached some of the representatives in the British Medical Association, general practitioners. You'll

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

see, GP throughout this, I'm afraid. GP is just the abbreviation for General Practitioner, roughly equivalent to your family practitioners. And they suggested that they could increase GPs incomes by up to a quarter, but they wanted that to be dependent on performance against a set of quality indicators. So between them, the Department of Health and some GP representatives came up with 146 quality indicators. These cover clinical care for ten chronic diseases, including coronary heart disease, [inaudible] diabetes, hypertension, mental health, and cancer. But they're also other indicators covering the way the care was organized, and patients' experiences of care.

And the way the indicators work is this. Each of the indicators is worth between 1 and 56 points, so in total, there's 1,050 points on offer to the practice. And each of those points will earn the practice the equivalent in year one, the equivalent of about \$145. Now that was adjusted upwards or downwards depending on the size of the practice population. So, for an average size practice, you're talking about \$150,000 per year available per practice, which works out at about \$47,500 per GP is an average of three and a bit GPs per practice. This is all additional money, so this is all on top of what the practices were already earning on a capitation basis.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

Just to give you an example of one of the indicators, this is one of the coronary heart disease indicators. So, the percentage of patients with coronary heart diseases, blood pressure is 150 over 90 or less. So the practice, if it achieves that target for at least a quarter of its patients will receive one point. So it's a minimum threshold that you have to get above to get any points at all. Points are then allocated on a sliding scale up to a maximum of 19 points available for achieving a target for 70% of patients. So you don't get any additional points for hitting a target of 80%, 90%, or 100% of patients. Where that maximum threshold is varies according to the indicator. So, on that indicator, the practice stands to earn anywhere between zero and \$2,750.

There's an important mechanism that was introduced by the British Medical Association, because it felt that if you were just blindly chasing quality targets, those targets might not always be appropriate. So, for example, you wouldn't want to be aggressively chasing the blood pressure or the cholesterol level of somebody who was totally ill, and just in order to get a financial reward. So the contract allowed GPs to exception report inappropriate patients. So patients who repeatedly failed to attend the practice, despite being invited, who got a terminal illness or are extremely frail or that couldn't tolerate, or didn't agree to a particular

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

investigational treatment. So you can see that it's not only protecting the GPs in here as well as protecting the patients.

And the way the system works is this. Going back to the coronary heart disease indicator I mentioned before, let's suppose that you got 100 patients on your coronary heart disease register. And of those, 50 of them have blood pressure below 150 over 90. Your achievement there would be 50/100, or 50%. So the practice would therefore earn \$1,530 for that particular indicator. But if, for whatever reason, ten patients were inappropriate, you could then exception report, and then those ten patients come out to the denominator, so your achievement is recalculated as 50 over 90, or 56%. So the practice would then earn \$1,870. So, no extra achievement, in terms of number of patients treated, but your effect of the achievement goes up.

The more devious among you will have automatically spotted a loophole. There's nothing to stop you inappropriately exception reporting, say, another 20 patients. If you do that, those patients also come out to the denominator. So your achievement, with no additional effort on your part, would then increase to 71%. You're above the maximum threshold, so you bring in the maximum \$2,750 into the practice. Now, that sort of behavior is referred to in the literature quite coyly as "gaming," and that's because,

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

particularly in the United Kingdom, we're far too polite to use the word "fraud." So, we were quite keen to investigate some of the consequences, so I just want to talk about a couple of them. One is an intended consequence. Was there a high level of performance on the clinical indicator? That, after all, is the entire purpose of pay-for-performance program. And I wanted to look at just one of the unintended consequences. What use practice has made of exception reporting, and whether there's any evidence that they used it to game the system, or defraud the system.

So, just as a summary, in year one, the first year of the program, which was 2004, 2005, the average level of achievement, the mean level of achievement, was 959 points, which is over 91% of the total. So a pretty high level of achievement, in terms of points. The average GP was earning an additional \$43,500. Remember this is on top of the salary they were already receiving. Two and a half percent of the practices managed to score the maximum 1,050 points.

Just looking specifically at the clinical indicators, the clinical indicators make up more than half the points. And there's an average performance of 530 points scored, which was over 96% of the total. So the practices were doing even better on the clinical indicators. And 6.5%, roughly, of the practices scored the maximum 550 points. But if you remember,

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

it's possible to score the maximum number of points without necessarily achieving all the targets for your patients. So what we wanted to do was actually look at achievement, which is what you can see on this graph here. So this is looking at year one. And this is looking at overall achievement for the clinical indicators.

So along the x-axis, you've got percentage achievement. So the percentage of patients with a particular chronic condition the practice achieved the targets for, and the number of practices along the y-axis. To give you an indication of scale, this is the 8,500 practices in England, specifically England. Which is 25,000 GPs providing care for just over 52 million patients. So quite a bold experiment.

The average achievement was 83.5%, but there was quite a wide range. They don't quite show up on this scale, but there were a handful of practices that got less than 10% achievement, and there was about 20, 25 practices that managed to get 100% achievement, i.e., that they claimed they hit every target for every patient with one of those ten chronic conditions.

Just to compare that with what the Department of Health have anticipated performance would be. We've got no baseline here, so we didn't know what a good level of achievement would be. But the Department of Health reckoned that the average

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

level of achievement would be around 70%, by the black lines indicated there. So you imagine they were quite surprised by the level of achievement in year one, and you can imagine the Treasury were even more surprised.

Going on to year two. Year one on this graph is indicated by the red curve, so in year two, you can see the average level of achievement, which was already very high, increase by over 3.5% in year two. I should say the financial incentives, if you thought they were generous in year one, they increased in year two by about 60%. The other thing to know about this distribution is you can see it's a lot more normal looking. So, the tail of really poorly performing practices from year one is starting to disappear, and the poorly performing practices got their act together a little bit more in year two.

Now moving on to exception reporting. Again, this is exception reporting on all the clinical indicators. So, along the x-axis, what you've got is the proportion of patients, percentage of patients on the clinical disease register, so the practice is exception reported. And again, the number of practices on the y-axis. So, on average, practices exception reported just under 6% of the patients on the chronic disease registers. Again, there was quite a wide spread, so there were several practices that didn't exception report any patients at

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

all. And again, they don't show up on this scale, but there were a few practices that exception reported over 80% of patients on their chronic disease registers. I'll let you draw your own conclusions.

Moving on to year two. You can see, again, year one is represented by the red curve on this. In year two, there was virtually no change of exception reporting, it went down very slightly. We were then interested to see if there was any relationship between exception reporting and achievement, to try and get at this issue of gaming. But we're aware that there are other factors you need to take into consideration. There are practice characteristics that might affect both exception reporting and achievement, including the ratio of GPs per patient where GPs qualified how old they were, where the practice was organized.

Some population characteristics, including the demographic structure of the population, the general health of the population. The socio-economic context, the level of deprivation in the area where the practice was situated, the level of unemployment, education, and housing characteristics. On the wider context, where about in the country the practices were located and which primary care trustee came on the - primary care trusts of the organization have oversight of the

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

practices and can go in and audit practices' data. And the payments to the practice come via the primary care trust.

So you can take all those factors into account if you have a seasonably sophisticated regression model. And I'll not bore you with the details, but just to say that these were the results. So, if you take each of those factors, which I've just outlined, you can represent them as a bar on this graph. So if the bar is pointing downwards, it means that particular factor is associated with a lower level of achievement. And if the bar is pointing upwards, it means it's associated with a higher level of achievement. So, for example, the second bar from the left, there is deprivation. So practices in deprived areas tended to have poorer levels of achievement than practices in more affluent areas. And some of the bars on the right hand side perhaps [inaudible] show the high ratio of GPs per patients tended to have high levels of achievement. It's important to point out that none of these factors have an enormous impact on achievement, but the factor which had the highest impact on achievement with exception reporting is represented by the big red bar at the end. And once you took all those other factors into account, a 1% increase in exception reporting resulted in an increase in achievement of about a third of a percent.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

There's a number of reasons why that might be, and I just want to cover a couple of them. The charitable interpretation is the practices that were good at identifying and treating patients with chronic conditions. They're also better identifying the hard to reach and difficult to treat patients who required to be exception reported. The less chance for interpretation is that these practices were exploiting the exception reporting system in order to gain, and then to increase their financial returns.

So, was there really any evidence of gaming? There were generally low levels of exception reporting in both years one and year two, and it suggests that if there is any gaming going on via exception reporting, that it was generally not very widespread, or not very large scale. However, there was a significant minority of practices, about 1% of them, excluded over 15% of their patients. And it was similarly our interpretation of these practices in particular required much closer examination, and through the primary care trusts, I recommended that they started auditing practices, at least [inaudible].

So, just to summarize some of the findings in the U.K. from the first year. The financial incentives were associated with a generally very high levels of achievement in the first year. But we have no solid baseline with which to compare

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

that, to see whether this was actually a reasonable level of achievement. Nevertheless, achievement did increase in year two, when the financial incentives increased by 64%. Whether that would have happened anyway without the financial incentives, we'll never know.

Population and practice characteristics certainly do have an effect on both exception reporting and achievement, but the effects generally were quite small. So unless you had a practice which had all of the characteristics associated with lower achievement, and all the ones associated with higher achievement, it didn't actually that great an impact. Our feeling is that large scale gaming wasn't widespread, at least using exception report as a mechanism to game. We might get on to this later, but there are 1,001 ways to game this system. We might be able to pick up on that later. But, certainly, using exception reporting, I don't think large scale gaming was widespread. You can get more details of the research we're doing from our website, I think you've got the details in your folders from me directly. If you want more details of the actual indicators, that website at the bottom will give you all the information you could ever want. So, I'll just turn it over to Bob.

[Applause]

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

ROBERT GALVIN, M.D.: Thank you. After the first two speakers and then someone with a British accent, you've just got this flat Ohio accent here, and it's all I've got. Greg set it up right. I only have a few slides, because what I want to do is get back to that table and have a dialogue with Tim. But I just want to spend the time on these slides to kind of frame the questions I'm going to ask, but what our country is facing and what we can learn from Tim. Unfortunately, we have a [inaudible] problem, but I will read it to you.

This slide is simply just to show you what I think Peggy showed earlier and talked about. There's tremendous activity going on in this country, but I think it reached a real, I think tipping point is the best term, really within the last year. The Leapfrog Companion's been around, that's got 100, 120 private sector kind of pilots and projects going on, on paper, performance. CMS has ten, actually ten demonstration projects that are all linking payment to performance. So, plenty of activity was going on. And then, I think two major things happened, and they happened about the same time. I only have one of them here. The first one is that the Institute of Medicine was asked by Congress through the MMA Bill to comment on pay-for-performance, among other things, and that report was co-chaired by Gail Wilensky and Bob Reischauer. I was in that

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

committee, along with several other people in this room, and our recommendations were published in August.

And I think it is worth going over them, because just remember, it's one thing for Medicare to have it in the demonstration. Much, much of what happens in demonstration projects never crosses over to the fee-for-service system, and that's what made this report important. And it was clearly the secretary should implement pay-for-performance in Medicare using a phased approach. Very clear recommendation. And let me see if I can even read this, I'll read it from here.

"The secretary should implement a monitoring and evaluation system of the Medicare P4P program in order to assess early experiences with implementation so timely, corrective actions can be taken." The font minimizes how important it was, in terms of the report. To acknowledge it, to move it to fee-for-service Medicare and the 40 million beneficiaries is a big deal. And the evidence base is weak, it seems to be intuitively correct, but we need to do this a little bit different than we've done other big changes before, in terms of this kind of learning system. Go to the next slide.

I should say before I go to this, the other thing that happened in August, it's important that Peggy did mention this, is about at the same time, the President came out with his

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

Executive Order essentially requiring that the federal employees who receive benefits, as with the federal government as a sponsor, meaning the federal employers through OPM, the DOD, the military employees, the VA, and several others, all were going to now be doing it in a system where the administrators were going to release data. They were going to kind of release data about performance, they were going to show data about price, they were going to do certain things around IT. And although no specific kind of requirement was made for payment, the fourth bullet of that is use of incentives.

So, I think you see a real movement toward this. The message is, "We're going to get going on this thing." So, I think now the key questions - and there are a million questions about pay-for-performance, I want to keep them at a very high level from a policy point of view, and I'm going to hit Tim with these - it simply doesn't work. This is going to be a major undertaking. And doesn't work. And I think you certainly, look at question number two, "Will the results be worthy ever?" Bruce Vladeck, one of the points he was making is, that information alone can do a lot. If you give providers information, you have the Don Berwicks, you have quality organizations working on it, a lot's done. Are all the resources that go into P4P going to be worth it? I think, third is clearly how we minimize unanticipated consequences and

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

I think we heard from many of the questioners, this is what happens in any forum that you bring this up. The many, many bad things that could happen if we don't do this right, and I think, most importantly, what can we learn from others. So, go to the next slide.

I did an interview with Tim's boss, Martin Roland, and it was published in Health Affairs six months or a year ago. And I asked him a lot of questions about their experience, Tim actually now has more data. And this is what I took it from. And this is very simple, but what we can learn, well, incentives clearly drove behavior that shouldn't have been a surprise to anyone.

But did you get a look at the extra money the primary care physicians were getting there? Those were whoppin' dollars. And I haven't done the complete math on it, but it would be, I think, somewhere around a \$10 or \$15 billion increase just in the Part B, as I was doing my math. Maybe David or someone over there can do it. That would be, in a budget-neutral environment like we have, a big deal.

I think secondly, it's pretty important. Know thy baseline. Tim didn't get into this directly, but you remember the slide where it showed what the government though was going to happen at the 78%? And then what happened in the overpayment. Well, they didn't' make that up out of nowhere,

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

it's a very organized, terrific system they have over there. And they took that based on what they knew about how the country was performing at the time. Well - and this is in my interview with Martin - it turns out that they had to do it at a certain time.

They made that decision, they start the program, and then further data comes out that shows that, in fact, the PCPs were already quite a bit better than the baseline they had suggested. And, although he used some very nice terms, "gaming," we would say, "We just paid for stuff we didn't have to pay for." And so I think, know thy baseline is very important. And I think the trust would verify he talked well about gaming.

We weren't really helped, because their system is different from ours on some key issues, and I'm going to talk more about them on my last slide in a second. They really didn't bring efficiency into it. So remember, they had made a decision to raise their percent of health care to GDP. They already were going to spend more money on health care, and they thoughtfully made sure it was for better care. We're in a very different situation, so there really wasn't a lot to learn. The minimally had some patient experience of care, although again, in the interview, the British physicians were not quite ready to do what we're doing with some of our patient

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

experience of care. So there really isn't a lot to teach us about how to integrate that.

We're very focused on silos, I think, appropriately that, you're in the hospital, you go home and see your doctor, there's frequently little communication and the many other kind of issues that happen. You go to see the specialist and the MR scan that you had isn't there, and there's really no bridging of selves [misspelled?], there really was no attempt there. This was a PCP-focused initiative for them. Go to the next slide.

So theirs is how we differ, and these are going to lead to some of my questions with Tim, which is, we need to do this on a budget-neutral basis, particularly if this new Congress does what it says it's going to do, which is reinstitute this PAGO, which is pay as you go, meaning - and they broadened it to mean we're not going to have new expenditures unless we find somewhere to take the money out, and that is something we're going to have to figure out.

Secondly, the IOM guidance was very clear. Which is, in a sense, don't do what they did. And don't do what we're doing now. Don't just do effectiveness and quality. You have to do efficiency and patient experience at the same time, or you're going to end up going down a road - and you know in

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

these big programs it's very hard to wind back. So, we're different, we need to figure it out.

We also believe that we need to preferentially reward IT. Now, in the U.K., and particularly in England, the government took it on, and basically bought IT for everyone. To the extent that it's going to cost over \$100 billion conservatively to wire our system, and these rewards are going to want to be directed towards IT, it changes our balance a little bit.

And then, fourth, we're very focused, I think rightfully so, on coordination of care. And I think as we think about not only doing effectiveness, efficiency, and patient experience, the idea, even on all of those, how do you coordinate across silos is going to be something unique for us.

So I am going to just stop. I'm going to walk over there, and put on my Charlie Rose mask. I don't want to stop audience questions, I saw the time there. But if we go question by question, I think that I asked Richard to solicit questions from the audience, and he said that the theme of most of the questions were already included in what I was going to ask. So let me ask a couple of questions, but I really do hope we have some time, because this is interactive, as Peggy said earlier. So... nice job, Tim. It always sounds smarter when you do it in your accent than when you do it in mine from Ohio.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

But I would say, ask, is the feeling in the U.K. - because I know there has been a little bit of controversy about the spending - that all things considered, this was good policy, and this was worth what it took to get there?

TIM DORAN, M.D.: It depends who you ask. I think the Department of Health are quite pleased with the levels of achievement. Not in terms of what it cost them, but they can then turn around and say that level of the quality of care being provided by primary care is quite high. And they can say that it was at least in part because of this program. I think the cost of it has taken a lot of people by surprise, and whether you could have got the same level of performance or a comparable level of performance for a lower cost, I think is exercising a lot of people.

There's also the issue of what I've been talking about, is just the hard data. Essentially looking at the figures and seeing how many GPs have actually hit the particular targets and how hard they've hit the targets. And there is a feeling amongst a lot of GPs that it doesn't reward aspects of care that they think are very important, and may actually draw attention away from those aspects of care. And further, as I mentioned, the clinical indicators only cover ten chronic diseases in the first quality outcomes framework. Others have since been added for this year, but there is a danger that by

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

concentrating on those indicators and those diseases which are financially incentivized, you're going to take the eye off the ball of other diseases which are not financially incentivized and care for those are going to suffer as a result.

ROBERT GALVIN, M.D.: I think that's right. Let me follow up on that, because that's a concern of everyone who's thought about it here, managing to the measure. And so, because you're ahead of us by several years, do you think it's addressable? In other words, have you thought about how to address it? And if so, anything you can share with us?

TIM DORAN, M.D.: Well, I think I'm going to sort of hedge here and say that it's far too early for us to say whether it's going to have, in terms of what's it going to do for aspects of care which aren't incentivized.

ROBERT GALVIN, M.D.: Well, that was question one and question two, is, have you thought of a fix to expand measures to the areas that you want?

TIM DORAN, M.D.: The thing to say about this framework is that the 140-odd indicators that were initially introduced are not necessarily fixed, and every two years they're revisited. And that's for a couple of reasons. One is to say that the whole framework is supposed to be evidence-based. There's no point in you trying to hit a quality target where there's no good evidence that it's going to improve quality of

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

care or going to improve outcomes. And so those would be revisited every couple of years. But also, the intention is to start taking away indicators which aren't seeming to be effective and to include indicators for other conditions that you think are important. So it has already evolved once, it will continue to evolve every couple of years. So, the idea is that we look at this sort of data and other analyses that we're doing at Center and other academic units are doing around the U.K., take those on board, and then keep developing the program.

ROBERT GALVIN, M.D.: So, let me bring you back to your hedge, which is, is there evidence about those areas that didn't get measured? Did you see - because that's a huge worry of people here.

TIM DORAN, M.D.: There's a technical difficulty with that, and you'll appreciate that most of the indicators that made it into the quality outcomes framework were things that could be measured. And if you're being cynical, it's things that GPs knew that they could do well. By definition, some of the other aspects of care are difficult to measure, and we don't have measures for them. But we are sort of looking back through other sources of data to try and measure aspects of care for other conditions and other indicators that were not incentivized, to see whether that's the case.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

Anecdotally, and talking to GPs, I know that they're saying that an awful lot of their work - you could say that only a quarter of their income is now dependent on hitting these indicators, but they get the rest of the money no matter what they do. So this is very much driving their activity. It's driving the nurses' activity. There is widespread concern that by focusing on this, by trying to get the last percent to get you over the maximum threshold it's taking an inordinate amount of time, and those other aspects of care are suffering. And in addition to that, there's a danger that care is going to get fragmented, so that you're going to sort of turn patients into conditions and conditions with indicators, and that's all that's going to matter to you.

So when somebody comes in, you're looking at your computer screen which is flashing up saying, "Ask them about smoking. Measure their blood pressure," where they could be dissolving in tears because they've got depression. So all those things have been coming out in the GP press sort of anecdotally, but no, we don't have any hard data yet to say that it's a large scale problem.

RICHARD GALVIN, M.D.: And I'm not exactly sure what to do about it. It leads me to a next question that I wanted to ask you, which is about the public. I mean, this is complicated, it's very much between the payer, the NHS, the

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

BMA, and the practicing physicians. Does the public know about it? What do they think about it?

TIM DORAN, M.D.: Generally, not a clue. [Laughter]
The minutiae of how GPs are paid is largely of little interest to the public. When we've done surveys on this, the public are not even aware that GPs are sort of independent contractors, are essentially running little independent businesses. They think that they're paid along - for those of you who are unfamiliar with the way that the National Health Service works, most people are given a fixed salary, and that includes doctors and secondary care. But the GPs are paid in sort of a business.

You're given a set amount of money for the number of patients you have, and you can spend that as you see fit. On nurses, on admin staff, on buildings, on clinics, or whatever. So all that is something that the general public generally are not aware of. What brought it to the general public's attention is that these additional payments to GPs increased the average salaries of GPs to above the 100,000 pound mark. So about \$200,000 has been average income for GPs in the U.K., and that was quite a psychological threshold.

They felt that anyone should be earning \$100,000, and in particular, that somebody working in the public sector should be earning 100,000 pounds took people aback. And then

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

they started asking questions about, "Well, exactly how is it that they're earning 100 -" So it's more the sums involved rather than what they were getting.

But going back to something Peggy mentioned earlier. There was a feeling, when details of the contract came out, of saying, we're paying this additional money to do things which are good for patients. So the question is then, what the hell were you doing before?

ROBERT GALVIN, M.D.: I think that's true. I can see that being a threshold. It's a good thing you decided not to give them stock options. Let me ask you kind of a hard question, which is, looking back now - this will be helpful to all of us who are really embarking on this - what's the one thing you wish you could undo? Or, even the couple things that you, you know how when you do one of these things you try, you always do your best to get it right. But in retrospect, there's almost always something.

TIM DORAN, M.D.: Not pay them as much. As has been mentioned, Martin Roland was involved in actually framing the indicators, coming up with the indicators. But had nothing to do with the sums of money that were not attached to them. And we've never been able to find anybody, and the Department of Health has put their head above the parapet and said, "Yes, I

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

fixed it at 76 pounds per indicator. And I decided to increase it to 126 in year two." Nobody will admit -

RICHARD GALVIN, M.D.: What a surprise. It's not like that in our country. [Laughter]

TIM DORAN, M.D.: I know. We do have a Freedom of Information Act, which we're actually thinking of using to go through some of the notes and finding out -

RICHARD GALVIN, M.D.: That would be very American, we would do that one.

TIM DORAN, M.D.: But yes. I think what happened was they did this backwards. They said, "We've got 1.8 billion pounds to spend. If they get the average level of achievement, what would that work out pounds per point?" and worked it out that way. Because if you do that sort of backwards analysis, you get to 76 pounds. Rather than saying, "What sum of money would you need to give to make these indicators effective?" And that's part of the problem that we have. It wasn't pilots, and it wasn't -

RICHARD GALVIN, M.D.: So you don't really know now.

TIM DORAN, M.D.: So we don't know.

RICHARD GALVIN, M.D.: Interestingly, in the interview, I asked Martin why they didn't pilot, and he said in our political system, by the time you do a pilot and think about it, it's like a demonstration project. You can get it going...

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

The Prime Minister's in place. And so to get the credit, and to drive the change you want to change with your limited time, you had to take the jump.

TIM DORAN, M.D.: Prime Ministers not so much, because this one has sort of hung on for the last nine years. It's the Secretary of State for Health, really. That is more of a hot seat. So no Secretary of State has stayed in post for more than two years since 1997. So if you're implementing something, you know, a couple of years down the line, if it works, somebody else is going to take credit for it.

RICHARD GALVIN, M.D. Let me do a process check. I've got one minute. Are we going to stay on that? Because I have the last question. Or we can go to the audience. So if anyone has any questions, walk up to the microphone now. No hand mics. And I'm just going to ask you one in case they don't. So now we need advice from you, and I'll tell you what your boss said after you tell me this, because I asked him the same question. [Laughter] You've shown us that incentives work. If you were here, and you had to do this program but you didn't have a billion eight. You somehow had to keep it budget neutral, or at least budget conscious. What would you do different? How would you go about it?

TIM DORAN, M.D.: Good luck. [Laughter]

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

RICHARD GALVIN, M.D.: That's what he said. He actually said, "I don't have a clue," is what Martin said.

TIM DORAN, M.D.: You'd have to do it on a much smaller scale. I think the 146 indicators would have been pared down to a much smaller set of indicators the physicians felt were both sort of evidence-based, based on quality, but also that they had a lot of control over. Just to give you an example. In the latest iteration, obesity has come into the framework. So, he hasn't registered this time, but there will future targets about getting levels of obesity down. And there's a lot of resistant stuff, because the physicians will say, "What control do we actually have over that? It's not fair to incentivize against that?" So I think you'd have the pared down version. You'd actually have to have much smaller payments to begin with. Because what you're talking about is winners and losers. Somebody's going to get paid more, that means someone else has got to be paid less.

The British Medical Association have got a good track record for digging their heels in for any sort of change, but they would never have gone along with anything which would have resulted in any GPs losing out substantially. And I imagine the situation's similar here. You're not going to have too many people saying, even if they're good performing physicians, "Yes, I'm quite happy with this system which means I'm going to

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

do better if I perform well but somebody else down the road is going to lose out." So I think you'd have to introduce it with much smaller size of financial incentives, and with a much reduced set of indicators.

ROBERT GALVIN, M.D.: So the summary of the advice is, better us having to do this than you, is what it sounds like. Adams?

MALE SPEAKER 1: Just to poke one more potential hole in this, the indicators are all interesting and important and things that a physician might have control over, like getting blood pressure down, getting cholesterol levels down. But the goals are easy and wouldn't be considered acceptable clinical goals in the United States. Getting the systolic blood pressure to 150 is not considered control of hypertension. Getting the cholesterol levels that you guys have used, I think they're not on your slide. They weren't translated to what we usually use. But I think it's a cholesterol of 240. And that wouldn't be considered cholesterol control in this country.

So, one way to reduce your payment is to, over time, move the clinical goals to what would be considered - if you had asked the British Medical Association, you didn't come from the NHS with a pot of money and you said, "Where should the cholesterol be?" they wouldn't have set the target that you set. So is that going to happen? And how do you make that

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

happen? It seems like right now that BMA has had an awful lot of influence over...

TIM DORAN, M.D.: Well, there's a couple things. One is, as I said, it's constantly under review. And so I think the initial thing is to say, "Well, let's get it above a certain level. Once we think we've cracked that, we can tighten it up." You can also tighten it up in another way, by increasing the maximum threshold, so for the coronary heart disease, why not put it on paper as a 70% threshold. You could increase that up to 80%, 90%. And they have done that in the latest iteration.

The other thing is, one of the diabetes indicators - I didn't put them up - for HbA1c levels, there's a two-step. So there's one set of incentives for getting it below 10 amongst diabetics, and there's another one for getting it below to 7.4. So you get a certain payment for getting it sort of moderately under control, and you get a higher payment for getting it very well under control.

So you could introduce that, you could have it stepped. But you're right, I think, particularly the levels of achievement that we've put up. Again, cynically, you could say GPs pick targets they knew they could hit, so let's tighten those targets. Now that we know they can do this, let's

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

tighten the targets for subsequent years and make them more difficult.

MALE SPEAKER 1: And is that solution [inaudible]?

TIM DORAN, M.D.: Yes. I think GP representatives accept reality. If you wrap it around 95 levels of achievement, you can't just continue incentivizing at the same level year after year. You have to make the targets more difficult.

ROBERT GALVIN, M.D.: Last question.

MALE SPEAKER 2: Thank you, Bob. Tim could you say something about how the GPs collected the information? I understand they all have computerized medical records, and the question is, did the computerized medical records support the generation of the measures? And if not, how did they generate the data to support the measures?

TIM DORAN, M.D.: The practices obviously involved, and the payment system all have computerized systems. And those systems differ. But they all kind of feed into the same general database. It was a fairly unique example in the United Kingdom of a large scale public sector IT program that actually worked, and you might have a similar experience in the United States. But essentially what it does is it piggybacks onto each practice's own IT system, and it draws the data out.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

So basically, if the clinician has entered, I have hit this target for this patient, this system then piggybacks onto that and draws that out onto a central database so that the GPs don't have to send the data, it's automatically extracted from their computerized systems. The incentive for you to then have a computerized system if you're paper-based is, obviously, the enormous sums of money that are attached to it. And the government will subsidize you to get a computer system installed in your practice. So, essentially, that's how the system works.

MALE SPEAKER 2: Thank you.

ROBERT GALVIN, M.D. Thank you. That was great, and thank you. And a real thanks to Peggy and NCQA for sponsoring this, on behalf of all of us. When I was on the IOM committee, the idea of a kind of learning culture in health care is something I think we benefit from. And I thought that was just a wonderful example of how we could learn a lot, so thank you.

[Applause]

GREG PAWLSON, M.D., M.P.H.: While we're changing teams here, I just want to make a quick comment. And I'm conflicted because I can't say much about this, but I'm aware of evidence that will be coming out at some point that is going to suggest, I think, that the pay-for-performance incentive did accelerate

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

the level of performance above the natural progression that was happening.

And the other thing is that at least in a couple of instances, it looks like indicators that weren't incented [misspelled?] may have remained at least on their same pathway of improving. And I think that's corroborated by evidence in this country within the VA system which has done a remarkable turnaround, in terms of incenting. And they actually incented at the practice group, or hospital level, that those things that they incented improve very fast.

They're now well above private hospitals in those parameters and the private sector in those parameters. And areas that weren't incented, at least didn't deteriorate, or remained in the same general relationship as they had with the private sector. And that was using the Rand Quality Toolkit stuff. And I think, maybe, some of our speakers may comment on that further.

GREG PAWLSON, M.D., M.P.H.: What do we know from the United States, in terms of all the various 100 or so programs that we've heard about that are in some stage in the United States, in terms of pay-for-performance? And what we tried to bring together here are the best person, in terms of policy from a clinical perspective in this area, the best health services researcher, and I would say the leader of a foundation

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

which I think has taken an incredibly thoughtful role in this area, who also is a veteran of health plan attempts to incent physician behavior and improve quality.

The people that we have on this particular panel are in your handout. They include David Blumenthal, who I've had the good fortune of being on a number of panels with. David is the Director of the Institute for Health Policy, and is a physician at the Massachusetts General Hospital, and has also been very active on the clinical realm, as well as being a member of the Institute of Medicine, and having been on virtually every policy panel that that organization has put forth looking at quality. Also has led a very important initiative for the Commonwealth Foundation that has also, I think, looked at a number of critical health policy issues.

The second person on the panel is also someone who I've been fortunate to be on programs with before, and that's R. Adams Dudley, who's an Associate Professor of Medicine and Health Policy at the University of California, San Francisco, who I think, unequivocally, has done some of the best health services research in the area of clinical performance, both on a hospital side and on the ambulatory care side, as well as a vast number of other issues related to quality and improving quality and health care.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

And finally, a friend of ours at NCQA, and someone who has been a friend of mine personally, as well as others at NCQA, is Steve Schoenbaum, who's Executive Vice President for Programs at The Commonwealth Fund. And who really brings to this, I think, an incredible perspective not only from his work at the foundation, but as I mentioned, as a leader before coming to the foundation of Harvard Pilgrim Health Care, having served as President of the New England branch of that organization. So Steve, you're going to start first, I believe.

STEPHEN SCHOENBAUM, M.D.: Thanks, Greg. I think we'll try to take you through these rather quickly, since we stand between all of you and lunch. And I believe we've set a 12:30 cutoff for this panel. We all have internal incentives, which is what drives us to do various things that we do, often our values and things like that. However, there are external incentives that the world tries to imply to apply to us that are very important. And those are things that basically motivate us to do things that we wouldn't have done under our own internal motivations. And I've listed several of them here. Peggy, this morning, showed you the way recognition programs can relate to several of those different types of external incentives, and much of the focus of today, of course, is on financial ones.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

I'm going to stay at still the incentive level for just a second. And sometimes people say, "Well, do incentives work?" that's a stupid question. [Laughter] Really, think about it. Because if it doesn't work, then it, by definition, wasn't an incentive. So it really comes down to how you structure incentives, as to whether or not they work better than other potential structures. And that means that the devil really is in the details, and the fact is that we don't have a lot of data at the moment, about how to structure the best financial incentives to achieve the kind of performance that we're trying to achieve in the U.S. health care system.

The recent IOM report was mentioned by Bob Galvin. He showed you the real cover of it. This might have been a better one, I think. And in that report, the committee that he was on reviewed the relative handful of studies that have been done in this area. There were nine randomized control trials, a total of 17 studies that they could find that they thought were worth reviewing.

And of the nine randomized control trials, four positive, two had a partial effect, three had no effect. And these are various kinds of pay-for-performance incentives for health care.

We know that at least a couple of years ago - this is data from a Commonwealth Fund survey back in 2003 of physicians

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

in the United States - that physicians said that incentives, financial incentives were a relatively small part of what determined their compensation and what motivated them to different behavior, at least ones that were based on quality. So we have at the top that productivity in billing was the major factor, not surprising.

That's the way the financial incentive system is structured. And at the bottom, about 19% said that quality bonuses, or incentive payments for insurance plans were either a major or minor factor in their compensation. Well, we all know that the situation has been changing, and that more and more people are being subjected or induced, if you will, with incentives.

So, here are some results from a 2006 survey. This survey was done among primary care physicians in the United State. It was done this past spring, and we see the number now is up to 30% of U.S. primary care physicians say that they have some kind of financial incentive that's related to quality of care.

And as you can also see, that's the lowest percentage of physician responding in the seven countries that are shown on this slide. And the highest, not surprisingly, is in the U.K., where you've just heard about the program. And if one looks a little at the details of that and just compares the

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

last two bars, which are U.K. versus U.S. performance, then most U.K. docs, at least the ones who were conscious while responding to this survey, said that they received a financial incentive for achieving certain clinical care targets, less than a quarter of U.S. physicians did.

When it comes down to managing patients with chronic disease and complex needs, it's only 8% of U.S. physicians who say they have such a financial incentive. These are, again, primary care physicians. So, while we're talking a lot about pay-for-performance, it's not clear that we've yet applied it to the degree that we might.

I would say as an introductory conclusion, that it's clear that we have a number of pay-for-performance programs underway, it's clear that we don't know as much as we need to know about them, and so it's clear, to me, that lots more evaluation is going to be needed. Because if you think back to it, the devil really is in those details, and just what those incentives were. Thanks.

R. ADAMS DUDLEY, M.D., M.B.A: Thanks for having me, Greg. I'm going to talk a little bit, a very little bit, about the evidence available. And in all the evidence that I provide, I insist that there be a control group. Because it's very important to know that the incentive really was something different than otherwise would have been. Steve started

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

talking about external incentives in our motivation, and I think it's possible if I had been born with a large trust fund, and didn't have the incentive to make myself into something other than that, I would get up every morning - I would get up.

Then I would go play sports, and then I would come home and take a shower and eat lunch and turn on ESPN, and before you knew it the night would be over. And then I'd go to bed again. And I don't actually do that, and there's some potential benefits to that. I have a wife and wonderful children that I probably wouldn't have if I did that. So we can't do that randomized trial and know whether I really needed the incentives to be at least a little bit productive, or whether that was really intrinsic to me without some sort of control group.

So everything I'm going to talk to you about here has a control group. Here's where you can go to get most of the available evidence about whether or not pay-for-performance works, but I'm going to talk to you in the next slide about - that it works. Pay-for-performance does work. You saw in the U.K. that it works. Of course, a hemoglobin A1c of 10, that's high. That's really bad control. It may not have taken that much to get people down to 10. But here's a study where it worked with a control group. At Vanderbilt, they randomized their residents to receive either \$2 a visit scheduled for

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

certain types of visits that they considered important, or \$20 a month just for attending clinic.

And the thing that they were going to measure was, were you there for all the Well Chart Care visits that your panel of patients needed? And, were you there for what they call continuity visits, which is when somebody in your panel had an acute illness and needed to be evaluated. And they believe it helps if you've known them and seen them all along. And the answer is that yes, they did, for \$2. Which, in Vanderbilt at about that time, was a slice of pizza and half a Coke.

So you can imagine that if they had done the same study with cardiothoracic surgeons, perhaps it wouldn't work. Anyone who's visited a medical school, if you bring along a pizza, everyone will come see you, at least if they're medical students or residents. And that's the sort of thing that does motivate them. But then, there are a lot of other people that it might not motivate.

Or, alternatively, pay-for-performance doesn't work. Here's another randomized control trial, equally well designed, in which they got completely opposite answer. So, in Philadelphia, a very respected health services researcher randomized a capitated [misspelled?] medical groups to be eligible for bonuses of \$600 to \$1200 for increasing colon cancer screening. And then there was a control group with no

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

incentive. And there was no effect of that bonus offered. Anyone have any ideas why?

Well, part of it was it wasn't worth it. You could argue, "Well, if I have to do a bunch of colonoscopies, how quickly do I get past my \$600?" And Greg whispered, "capitated" up here. And that, to me, is the biggest, most obvious thing that was going to happen. So, colon screening, cancer screening, by definition, is going to asymptomatic people and asking the clinical question, "Does this person somehow have cancer?"

Well, if you're capitated, and the answer is, "Yes," then you've got to pay all their downstream costs. So this was one of Steve's non-incentives, right? Because, in fact, if they didn't find any cancer, then a bunch of people who probably weren't looking to have colonoscopies had them and didn't get any clinical benefit out of it, and if they did find any cancer, they lost money. So, this is a non-incentive.

So the devil is in the details, as Steve said, and so I'll just suggest what some of those details are. The magnitude of the financial incentive is important. You could also add, since you've collected the study, you could add reputational effects. I think providers care a lot about their reputation. I'm not so sure that they really believe that consumers are going to use it, but they're concerned that the

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

people they're going to meet in the staff room are going to look at the data and they don't want to look bad in front of their peers.

And then, the costs of complying are important. Other negative studies, we've heard about the obesity. Do providers in the U.K. think they can move obesity? Well, if they could actually improve that, you know that it's going to be a lot of work.

And so, the incentive that you need to get someone who is obese to actually lose meaningful weight and keep it off is probably greater than the incentive you need to get someone whose blood pressure is a little bit elevated down to the goal of 150 over 90. And so the cost and complying ought to be taken into account.

Other factors that might be important are the business environment itself. In the study we saw, a fee-for-service for capitation, if you had said, we'll give you a \$600 bonus for doing more colon cancer screening in a fee-for-service environment, that actually completely switches it around to an actual, positive financial incentive. So the overall general payment or business environment matters.

It probably also matters that the reason I use business instead of just payment, it's not the rest of what you're paying. In the U.K. it is, because they're the only payer.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

But if you're anybody in the United States, including even Medicare, it probably matters what everyone else is doing. And if somebody else is saying do one thing and you're saying do another, there can get to be what I term, "incentive cacophony," where it's just too many signals, and in the face of too many signals, people just ignore all of them. And so it's conceivable to create that.

You saw in the U.K. the characteristics of the provider mattered, that the providers over 50 actually did worse with the incentive system than the providers under 50. And that's a similar finding that can be made in the United States. Organizational characteristics also matter. So information technology allows people to respond to an incentive, for example, more easily, and so if you're going to a place where IT is widely available, you may be able to set higher targets.

Similarly, in California, for example, we have many large provider groups. Of course, we have our onesies and twosies providers, but you could go out and interact with a very large number of large provider groups who cover a very large percentage of the total population.

There are many states in the United States where that cannot be said. So it may be that what you can expect out of a group that has scale and resources is different than what you can expect out of a practice with two physicians. And patient

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

factors also matter. And that's been largely ignored, so I'd like to get to that in one second. This slide here, it's too hard to read, so just look at it later at your leisure. What I've tried to do is list all the details in Stephen's devil's slide, and if you're interested in designing a program or someone's trying to design one to which you have to respond, go through these and see whether or not you feel like the program addresses all of the things that need to be addressed.

I'd like to spend the rest of my time on the impact of patient factors on measured performance. I've heard over and over again that process measures like NCQA has in its HEDIS measures, or JCHO does, and what it's release is don't need to be risk adjusted. And I actually sort of believe that. It just is, that we want to get the hemoglobin A1c checked. And it just is that we want to get blood pressure below 150 over 90.

But it actually turns out that patient factors do matter, and so I don't know whether risk adjustment is the right thing. We'll talk about what's the right response. But let me just show you, try to convince you that patient factors do matter.

This is data from a study that we did with the CDC, which at one point defined managed care as a public health problem, and therefore funded us to study the impact of managed

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

care features. CDC actually funded a nationwide study to look at the impact of managed care features on quality of care in diabetes. But one of the interesting things that came of that is we got together 11,000 patients from health plans in six sites around the country who had diabetes, and it actually turned out that race ethnicity had a huge impact on your underlying health state that, if you then turn it around in to pay-for-performance you can see how it might matter.

So if we look here at the hemoglobin A1c measures over there, you see the racial groups vary from 8 to 8.5 in their mean hemoglobin A1c. Now, we heard in the U.K. of performance targets of 10 and 7.4. Most U.S. pay-for-performance programs use a hemoglobin A1c of 8. So, if you're looking either at that 8 or that 7.4 target, if you happen to have a lot of people from the purple racial ethnic group, you're starting a whole lot closer.

If your population with diabetes starts balanced right on the payment line, and even if you're doing 7.4, they're less than a point away. And if you've got a lot of people in the gray group, then you're more than a point away. This is just race ethnicity where we found a baseline. Similar points can be made for LDL and cholesterol goals that you're shooting for, and blood pressure goals as well.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

So, patient characteristics may matter, and they may matter in another way. You go out and do the best, so you've got your population, you've got your instructions about what the thresholds you're trying to hit are, and you do your best, and you go out and you write your prescriptions. So if it's hypertension, you've written a prescription for someone to take of a medicine that lowers their costs. Now, if you're in the United States, almost certain that the patient has to pay for part of that prescription, right? Very rarely these days that people get prescriptions for free.

Well, it turns out that people stopped using their medications because of cost. And that varies by a bunch of demographic factors, so it varies by race ethnicity, it varies by age, with younger people being more likely to do it. It varies by gender, and also, if you notice the middle column, then even though people do this with some frequency, they don't tell their doctor about it.

So if you're out there trying to meet your goals and you're furiously writing prescriptions, and you have a Latino population, on average, you'd have 19% of them not taking their prescriptions because of the cost, and then less than half of them telling you that they were not taking their prescriptions because of cost. It creates an exception reporting nightmare,

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

and makes it difficult to be sure that you're measuring people fairly.

I'll skip this next one. More factors that influence this. We also did a study in the same population where we looked at persistent lapses of what predicts that a patient will have missed the typical diabetes performance measures that are widely used that are in HEDIS for three years in a row. So for each measure it only counted if you were without it for three years in a row. And it turns out that younger aged, less severe diabetes, having fewer core morbidities [misspelled?], having a job, and being at a low income level all predict missing these interventions. Completely separate from anything that doctors do. So this is after correcting for which practice you're in, and whether they have a registry, et cetera.

And which health plan you're in and what the benefits are, et cetera. All these factors had big impact on whether or not you were - and the real tragedy here is these are the people where you actually want to have an impact. So getting the retinal screening after the patient's already begun to lose some of their vision isn't nearly as effective as keeping them from losing their vision in the first place.

And it's not just that it has the greatest clinical impact, it's also the most cost-effective. So as we're

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

proceeding forward with performance measurement and with pay-for-performance, we've got to be really careful about the performance measurement. And in particular, I said I'd talk solutions. I know that there was a suggestion for us to keep it simple, but I just don't think we can.

I think the reality is that it is worth more to do certain things than to do others, and it's worth more to do the same things on certain people than others. And it's harder to do the same thing on certain people than others. So I wouldn't actually risk adjust, because if you think about it, if we risk adjusted for these various things to try and achieve what we wanted, then we'd be saying, in the case of hemoglobin A1c, "Well, it's okay to have a worse hemoglobin A1c in people from certain ethnic backgrounds." And that's kind of a hard sell, or at least, I hope it is.

For these persistent lapses measure that I think is probably really important, it's really important who's just, over and over again, missing their marks, we'd end up saying, "Well, it's okay if they're older or sicker people," and you'd sort of get credit. And that wouldn't be great either. So, instead, what I would say is, let's make the payment more when it's harder.

So if for a certain ethnic group it's harder to get them down to 7.4, then the payment should be more for that.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

And if a person has a job and three kids, and they're a woman, and all those data say that's the person that's going to be really hard to get them to take care of themselves, then the payment can be more than for the slacker man I would have been, probably would have gotten diabetes, if I had had a trust fund.

It wouldn't have been that big deal for me to miss a game and go in and get my test done. So maybe that's a solution that we can come to down the road. That has not been done at all, anywhere, in any country. But I think that in the end, if we want to set the incentives and have them be effective, we have to reflect what's hard to do and what's easier to do, and what's important to do, and what's less important to do. And I will stop there.

DAVID BLUMENTHAL, M.D., M.P.P.: It's a pleasure to be here. I want to echo what Bob Galvin was beginning to say about the quality of this program. Peggy, and Richard, and NCQA deserve a hats off for a terrific program. I'm reminded a little bit of the time when a foundation was kind enough to fly me to Aspen for a meeting. It wasn't the Commonwealth fund, Steve.

And I was taking some time off to play tennis, and after I figured out that I needed a different ball so that my forehand didn't look like a Tiger Woods tee shot, I noticed that there was a woman playing next to me on the court who was

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

just really something different. She really played tennis well. And I went up to the office afterwards and asked who that lady was, and the woman looked at me like I was a dimwit and she said, "Oh, you mean Billie Jean King?" [Laughter] So ever since then, I've said I played tennis with Billie Jean King.

I feel like that today on this program. I don't know if Tim would call that gaming or what, but I'm going to take a somewhat different tack, and I don't have a lot of time.

Peggy opened the topic for me a little bit by saying that P4P paying-for-performance may not be enough, and in keeping with Newt Gingrich's billboard analogy, the point I want to make for you today is to think about local markets. Think about the role that local markets have to play, in my view, should play, in making pay-for-performance work. So, I'm going to see if I can advance through some slides here.

Why local markets? Local markets, I think, are the smallest meaningful level at which collective action and coordination can occur. And I'm going to make the case that pay-for-performance is going to be much less effective, here we're talking about the devil in the details, it's going to be much less effective unless it's coordinated to address the problems that are facing local markets.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

Local markets are where people get their health care. Thus, they define consumer's choices for alternative types of health care. I don't really believe that most of us are going to be getting on planes flying around the country getting our diagnostic workups, our colonoscopies and the other routine care we get. They define populations that can be the target of population-based treatment, and they define where things go wrong.

We know that from Jack Lindberg's work that there's a signature for every local market, and they may be the area where things can go right, as well. And I'm going to go through a few turned up diagrams here.

No health services researcher is complete without a turn up diagram. But just to use some data that I actually got from Steve Schoenbaum, and a slide that The Commonwealth Fund has used, this slide looks at adjusted one-year survival after acute myocardial infarction by hospital referral region, and also looking at cost of care.

And it strikes me that the pay-for-performance program you need for the hospital referral region on the far left may be different from the pay-for-performance you need for the hospital region on the far right, in terms of where you focus people's time and attention.

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

So, the counterfactual is what some of the audience members have raised. If each payer and regulator in a given market, a given location, has a different strategy for performance improvement, will their programs be effective? This is the chaotic, or cacophony question. Medicare, Medicaid, private payers, each have different pay-for-performance programs emphasizing different indicators.

Maybe even they have different measures for the same performance, and also different disease management strategies. Will the result be optimal performance at the local or regional level, or for that matter, at the individual level? And I think one of the few things that may be growing faster in our health care system - this is a sign of progress, I don't mean to diminish it - one of the few things that may be growing faster than the cost of care is the number of measures we have for quality.

At Partners Health System in 2000, we were reporting about 30 measures. We are now reporting 160 measures. I don't want to make analogies to the percentage of GDP that we project ahead, that we're going to be more than 100% of GDP for health care, but at this rate of growth, there does need to be some restraint on what it is that's requested.

And once you start paying for those measures, it becomes even more important. So what you call a local market

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

as a vehicle for activity and coordination depends a lot on where you are. It could be a state, as in Rhode Island, it could be a county in a rural area, it could be a city, as in Pittsburgh.

It could be part of a city, like, maybe Manhattan or even the southern part of Manhattan. But it's defined by the opportunity for collective action and for the identification of distinctive patterns of care and provision and receipt of care by providers and patients.

What are the implications for pay-for-performance? Thinking about this in eastern Massachusetts, which is where I go to work every day, how many different types of performance is my health care system and I personally, as a physician, going to be rewarded for paying attention to? How is that performance going to be measured? How large will the rewards be? How predictable - and this is critical - how predictable will the programs be over time? We love innovation in this country. The death of pay-for-performance will be constant changes in what's measured and in how much you're rewarded for.

If consistent answers can be found across payers, then the impact is likely to be much greater. I don't know if we have to pay increased salaries by 20%, but there is some number that will get people's attention so they won't drive by it like driving by that billboard. And if providers that contribute to

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

the development of these patterns of reward, they're likely to be much more effective.

One of the things that's nice if you focus on markets as an area where some of these issues can be worked out collectively, is that that creates the opportunity for cross-market learning. And I want to emphasize that pay-for-performance is a solution. We shouldn't lose track of what the problem is. The problem is performance.

Performance is going to yield to multiple simultaneous interventions that have the possibility of undermining one another unless they're coordinated in their impact. So, what are some of the other things we need to pay attention to? Lack of alignment between supply and demand in the local market. Local cultures of practice. We know that local cultures of practice vary enormously. They didn't get that way because the financial incentive is different, they got that way for other reasons, in many cases. There are varying levels of competition in different markets, and varying levels of supply. And there are varying levels of involvement, motivation, and stability of key purchasers, i.e., employers and health plans.

So, I didn't talk about macrosystems, but ignore that term for the moment. I think pay-for-performance should be incorporated into comprehensive, long-term, market level reforms could change the system of care within a market. I

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

think the strategy should be tailored to local markets, structured local markets in the pattern of variation. It should include both long-term and short-term goals, and it should include coordinated efforts at the evaluation of the varied strategies and combinations that each local market will imply. Thank you.

[Applause]

GREG PAWLSON, M.D., M.P.H.: My only real sadness here is that this panel has so much more that I know they know, and hopefully you can consult some of their publications for some follow-up. But I think they've each done a remarkable job in condensing knowledge into very major, hopefully bite-sized bullets. We have time for just a couple of questions. Peggy?

MARGARET O'KANE: I just have a question for David. Do you think Medicare ought to continue its pattern of paying 2.5 times in Miami what it pays in Minnesota in this local market scenario, or should Medicare be thinking about some kind of parity across markets if we were to move to that model? And I guess I would ask that question even if we didn't.

DAVID BLUMENTHAL, M.D., M.P.P.: What I'm thinking about is local markets as a target for reform. So, I don't think that Medicare can stop paying overnight more in one market than it can in another. But I do think it's fair for

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

Medicare to ask local markets to achieve changes in performance.

The question is whether Medicare should dictate the microlevel nature of the reforms, exactly what they pay for and how much they pay for it, or whether they should set - say, you have a toolkit for reform. Among them could be paying for performance. We will try to adapt our local behavior to the reform strategy that you have derived, and we'll hold you accountable for the performance improvements that you promised.

R. ADAMS DUDLEY, M.D., M.B.A.: To underscore David's local markets argument, Newt... are you still here? I don't think people are going to be jumping on planes. In fact, in a study done at Dartmouth of people who needed prostate cancer surgery, they asked them - admittedly the question was a little abstract - if there were a hospital that had a lower mortality rate, how much would you be willing to travel?

And the rough calculation people seemed to make was about three miles per percent in mortality. Of course there are counter examples. You hear of the person who gets cancer in Tampa and drops everything and goes and gets a bone marrow transplant at Duke and it's on 20/20 or whatever. But that's not most people, and it's not most illnesses.

The second thing is that it couldn't possibly happen that way, there just isn't sufficient supply. So what that

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

means is that we have to organize at the local level, but it puts an intense pressure on picking the right partners and making sure that the partners are there at the local level. And that can vary, depending on what that local market really looks like.

GREG PAWLSON, M.D., M.P.H.: I think one of those ones [misspelled?] we ought to be aware, or all of you ought to be aware of, some of the more successful and emerging regional projects, like the Massachusetts Health Quality Project, the Rigen Street [misspelled?] Indianapolis Project, Puget Sound, Minneapolis, and the Wisconsin Projects, as well as IHA and PBGH in California.

And they're really beginning, I think, to emerge and to be able to combine data across health plans, across medical groups and really begin to get over the problems with small numbers and some other things that have plagued us. And also to really meaningfully connect up with the medical groups in those regions. So I think the suggestion of the regional stuff is really terrific. One more question.

DOUG HUFF: Doug Huff, Johns Hopkins University. Something that Dr. Blumenthal mentioned that sort of triggered something on me. In P4P, are we talking about pay-for-performance, or pay-for-improvement?

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

GREG PAWLSON, M.D., M.P.H.: Very good question.

Adams, or David?

DAVID BLUMENTHAL, M.D., M.P.P.: Well, I have a prejudice about this, though I realize the problem. I much prefer paying for improvement. Perhaps subject to some minimum threshold of performance before the payments kick in. So, the problem is always physicians performing well or plans performing well getting penalized because they can't improve as readily from a high level of performance as you can from a low level. And it doesn't seem right to reward low performing people more readily, or groups.

But I think once you get to a certain minimum level, then I think improvement is really what we should be rewarding, rather than just performance.

DOUG HUFF: So the reality of payments so far is that it's mostly gone to people who are already performing well. Not just in the U.K. So PacifiCare had a very well intentioned P4P program in California, and that was studied by Meredith Rosenthal with the comparison group being PacifiCare groups in the Pacific Northwest. And it turned out that there wasn't that much improvement overall in California relative to the Northwest. And then when you look back and ask Steve is it really an incentive question, the answer was mostly not. Because the large majority of the money with the incentive was

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

set up that you get a bonus for going over the 75th percentile of last year's performance.

Well, most of the money went to the people who were already above, and they didn't have to do anything to get above the 75th percentile. And then, of the people who weren't at 75th percentile last year, the people in the bottom 50% probably thought there was no chance they could get there. So you were really, basically only giving incentive to that small subset that was near the threshold.

Now, you didn't have to set it up that way, you could have set it up in a way that would have given more incentives to improve to more people. It's just that we have to be very careful about how we do that.

And then one other thing I think we ought to think about, is in our choice of measures, this concept hasn't gotten out. The measure as archetype. A measure as reflecting something that, if we improved on that, it would help in more than just the measured area. And by this, I'll use the example of two measures that we frequently use that are timeliness [misspelled?].

One is use of antibiotics in pneumonia, and the other is time to PCI in MI. Fixing the time to getting PCI is getting a stent placed. Fixing the time to getting your stent

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

in MI is a one-off intervention. You fix that, you fix nothing else, I think.

Fixing time to antibiotics in pneumonia applies also to other situations where you need antibiotics. I was just in the hospital for meningitis, would have been nice if I had gotten my antibiotics a little faster. And also, to anything else where you need to mix a drug at the time, so it would apply to getting intravenous steroids in asthma.

And so, as we pick our measures, we perhaps ought to be paying less attention to what's feasible now and more to what would reflect a major change that matters over the long term for more than just the one area where we're measuring.

GREG PAWLSON, M.D., M.P.H.: I would also be remiss if I didn't plug the potential of using systems measures as a means of pay-for-performance, because I think there's some reasonably strong evidence that improving systems, like getting a registry, identifying patients who aren't doing well, et cetera, can actually lift performance and a whole bunch of bad measures.

Thank you very much to the panel.

[Applause]

RICHARD SORIAN: I want to thank our morning's speakers and panelists. I want to apologize to those whose discussions

¹ kaisernetwork.org makes every effort to ensure the accuracy of written transcripts, but due to the nature of transcribing recorded material and the deadlines involved, they may contain errors or incomplete content. We apologize for any inaccuracies.

were compressed a bit. It's the way conferences often go, but we do appreciate the tremendous time they put in for this.

We're going to go to lunch now. Lunch is at the Atrium Ballroom, which is on the mezzanine level, which you can get to either by stair or by elevator. Leslie Norwalk will be speaking at approximately 1:00, so we should be able to get back on time. And we're coming back here after the lunch for Congressman Stark and for the Thought Leader Roundtable. Thank you very much.

[END RECORDING]