Session date: 11/16/2015

Series: Evidence Synthesis Program

Session title: Understanding the Intervention and Implementation Factors Associated with the Benefits and Harms of Pay for Performance Programs in Healthcare

Presenter: Devan Kansagara

*This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm.*

Molly: And we are at the top of the hour. We do have a nice big panel presenting for us today so I am going to get through these introductions as quickly as possible. Presenting the bulk of the report findings will be Dr. Devan Kansagara. He is a staff physician and investigator at the HSR&D Center to improve Veteran involvement in care and with the ESP Program located at the Portland VA Medical Center. He is also an Associate Professor of Medicine, Division of General Internal Medicine at Oregon Health and Science University. Joining him to present some of the results will be Aaron Mendelson; he is a research assistant at the Center for Health Systems Effectiveness at Oregon Health and Science University in Portland, Oregon. Joining us but unable to speak for a large portion of it today is Dr. Karli Kondo, she is an investigator at the VA ESP Program Center in Portland and in HRQ Scientific Resource Center and Assistant Professor at Public Health and Preventative Medicine at the Oregon Health and Science University/Portland State University School of Public Health. Also joining today is Dr. Cheryl Damberg, she also helped with the report findings and she is a RAND Distinguished Chair in Healthcare Payment Policy and Principle Senior Researcher at RAND Corporation. Finally providing comments after the results have been presented will be Dr. David Atkins, he is the Director of Health Services Research and Development for the Department of VA and also Dr. Joe Francis the Director of Clinical Analytics and Reporting in the Office of Analytics and Business Intelligence for the VA, Veterans Health Administration. With that Devan I would like to turn it over to you if you are ready.

Dr. Devan Kansagara: Okay thank you Molly. As Molly had mentioned, these are some of the names that are acknowledged here. I just want to make a quick acknowledgement – I am not Karli Kondo but she is the lead on this project and unfortunately had fallen ill last week and cannot talk for too long without coughing. I am presenting in her place and I will try to do it justice. This report was a lot of work and she did a great job leading a large overview of the evidence which we will get into here.

Listed on the right are the people that nominated the report are stakeholders and peer reviewers so I will just jump in. The first few slides here are just standard slides about the ESP Program that I will skip through quickly and that you can go back to if you want to learn more about the Evidence-Based Synthesis Program which does systematic review work for a variety of VA stakeholders.

On this slide, at the bottom I will just point out a link for those interested in nominating topics for future years, there is a link and some instructions there. The link on the bottom of this next slide is the VA ESP link and our reports are posted there. To get into the current report, which I will just mention is on the internet currently and will be released publicly in a few days and the link to the VA ESP site and the report is there.

Quick overview of today’s presentation, we will do a very brief background and then talk about the methods guiding this systematic review; talk about the results of three key questions. Then we will pause and hope to get through all of that in about twenty-five minutes so this is really going to be a pretty brief overview of the results and it is a large report so there are a lot more details there if people are interested in reading them. Then we are going to pause and hand it over to Dr. Atkins who will discuss briefly the implications of these findings and how it relates to future research efforts in the VA. Then we will be joined by Dr. Joe Francis who will talk about the policy implications in the VA. Then we hope to leave plenty of time for discussion and Q&A.

To jump in here, as I am sure most everybody is aware the healthcare environment has changed dramatically in the last decade and pay for performance programs have been a big part of the changing landscape. They have been implemented in a variety of settings including the VA with the hopes of improving efficiency and quality of care. There have been a number of recent reviews attempting to summarize the literature and many have found insufficient evidence to broadly characterize the balance of benefits and harms. In particular we will talk about the RAND Report when I get to the methods here as one of those recent reviews. However Pay for Performance Programs are really complex, incredibly complex interventions and their effects may depend in part on the settings in which they are implemented; the methods used for implementation; the patient populations that are targeted and other program characteristics. That was a big motivating issue that underlies the nomination of this particular report.

The scope of the review was based on three key questions. The first is – what are the effects of Pay for Performance Programs and patient outcomes and processes of care? Aaron Mendelson is going to discuss those results in a moment. The second question has to do with - what implementation factors modify the effectiveness of pay for performance? And the third key question centers around - the positive and negative unintended consequences associated with pay for performance.

The inclusion and exclusion criteria guiding the reviewer listed here, the P-codes are listed here so the population we are interests in was broad, we did not place a lot of limits and we were looking at pay for performance programs that directed incentives at either the individual provider level or the managerial level, group level or institutional levels. We are focused obviously on financial pay for performance programs, we did not look at broader systems changes that may overlap some pay for performance such as accountable care organizations and bundled payment pilot programs, and those are not included in our review. The outcomes of interest were also fairly broadly defined and included intermediate and final health outcome measures for patients and process of care measures. The settings obviously included the VA, other healthcare systems in the U.S. and of note we decided to include evidence from other countries because we thought that there were instructive examples from other countries such as the U.K. who have elements in common with the VA healthcare system.

A little bit more on the methods, our search strategies which are provided in detail in the report were based largely on the RAND Report to which I previously referred. So I will just pause here for a moment and say that Cheryl Damberg who is here with us on the phone was the lead on that report and is an excellent very large comprehensive report focused on U.S. pay for performance programs. So we updated their review and then we also went back in time for studies in other countries such as the U.K. We searched the usual databases and additionally we searched the gray literature and websites for additional information. And we also did some targeted searches for specific pay for performance programs that we had known about including things like the quality and outcomes framework in Britain.

In terms of the types of studies we included, we used a “best evidence” approach so we focused on larger studies that included a comparison group or that did not include a comparison group but that had at least three time points analyzing trends for KQ1. Our inclusion on study design was much broader for key questions two and three that focused on implementation factors and unintended consequences. As I mentioned we were fortunate to work with Cheryl as a co-author on this review so she was really able to work with us on summarizing the RAND Report findings as they pertain to the VA and kind of pulling out the best quality information from that report. In addition because a lot of this was focused on implementation processes and unintended consequences we felt there were probably things that were not included in published literature so we conducted key informant interviews with experienced pay for performance researchers; we conduced fourteen interviews each about an hour long and we took live notes during the interviews; distributed the live notes to our five investigators who independently reviewed the notes and identified themes from these notes and we consolidated those and will describe both the results of the literature review as well as key informant findings throughout the rest of the discussion. Mostly the key informant findings will be relevant to key questions two and three.

Our search yield we searched thirteen hundred citations; five hundred full-text papers and of these we included ninety-three across the three key questions. Some of the studies answered more than one key question and as I mentioned we conduced fourteen key informant interviews. These were studies in addition to the studies that were included in the RAND Report.

With that we are going to move on to key question one which looks at the effect of pay for performance programs on patient outcomes and processes of care outcomes. So take it away Aaron.

Aaron Mendelson: Alright you want to handle the clicking for me.

Dr. Devan Kansagara: Sure.

Aaron Mendelson: Aaron Mendelson here, I am going to be presenting from KQ1. For KQ1 we are going to lay out the findings by outcome and then by place of service and [00:12:07] [background noise] studies that exam the process of care. Of those thirty-six examined in the ambulatory setting, we found modest improvements associated with U.K.’s Quality and Outcomes Framework with the largest improvements in years one and two, followed by a plateau or slowing of improvement rates thereafter. For example a study by Doran and others in 2011 examined twenty-three incentivized indicators over a seven year period. There was a four year low \_\_\_\_\_ [00:12:33] [audio skipped] pre-intervention, all seventeen process of care indicators improved significantly in the first year. The third year achievement \_\_\_\_\_ [00:12:42] seventeen indicators remain significantly higher than projected. However between the first and third year achievement plateaued with mean rates increasing by only 1.9%. For the non-QOF studies there were mixed findings in the U.S. and in other countries. Some studies reported modest short-term improvement such as Taiwan’s diabetes pay for performance program and others particularly longer-term studies reported a slowing of improvement or little to no effect.

For example five studies reported findings from Taiwan’s Program. In 2001 as a voluntary program focused on guidelines \_\_\_\_\_ [00:13:20] [audio skipped] continuing medical education training, participate, pay for performance was significantly associated with increased screening rates and survival. But physicians who had completed the required CME but chose not to participate in the program also screened patients at a significantly higher rate than physicians who are program ineligible. [Next slide please].

So we included six studies evaluating process of care measures in a hospital setting. The Premier Demonstration and the Hospital Value-Based Purchasing Program showed no significant improvements. However, a study of a VHA program targeting acute coronary syndrome, heart failure, and pneumonia process measures reported significant improvement for six of the seven measures examined – international studies generally reported positive effects, with again a slowing of improvement or a plateau over time. [Next slide please]

So we included twenty-three studies that looked at patient outcomes and of those nineteen were looked at the ambulatory care setting. Overall there was no clear evidence that the QOF increase increases clinical target achievement, achievement for some managers such as HbA1c were actually lower than the pre-QOF trend would have predicted. And in areas that improved, the greatest improvement was typically in the first year and again, there was a plateau or a slowing of improvement thereafter. One good study that illustrates this was by Vamos [ph] and others they looked from 1997 to 2005 and they found that there was immediate increase in blood pressure achievement with another increase the year after. However for cholesterol they did not find any immediate effects with a slight improvement the year after. For HbA1c there was no immediate improvement and then the following year there was actually a non-significant decline in target achievements. Then slide seventeen.

There is little to no evidence in the U.S. and Taiwan that P4P increases clinical target achievement with Taiwan’s diabetes program associated with no significant short term effects but marginally fewer diabetes related complications and hospitalizations in the long term. Studies in the U.S. reported fewer E.D. visits but marginally higher acute and ambulatory care sensitive hospital admissions. [next slide please].

Then of the twenty-three patient outcome studies, four of them examine the hospital settings. Studies in Taiwan reported higher five year breast cancer survival and lower recurrence rates as well as higher tuberculosis cure rates. There were no improvements in patient experience associated with the HVBP. And the U.K.’s HQID was associated with short term improvement with controlled studies showing no effect in the long term. So a 2014 study by Christenson compared P4P and non-P4P hospitals and measures of risk adjusted mortality for patients with incentivized conditions related to AMI, heart failure and pneumonia. They found risk adjusted mortality decreased for all eight conditions in study and control hospitals as well as all of England during the study period. While the intervention did have a significant short term effect; mortality rates for non-incentivized conditions and incentivized conditions showed similar or larger decreases during the time period. Short term improvements were not maintained and there were no significant differences between HQID and control hospitals before and after the intervention. That is it.

Dr. Devan Kansagara: Okay thank you Aaron. We are going to switch gears now and get into the implementation processes and their impacts on outcomes. Key question two was guided, our organization of it was guided in part by this framework that is on the slide here. For time sake, we did not include our conceptual framework but I will just quickly note that the conceptual framework guiding the report was based in large part on the CEFR model and others work including Laura Damschroder and Cheryl Damberg’s work on the RAND Report.

I will just quickly go through this slide so we are oriented on the terminology we are using. Program design features were really the properties of the intervention themselves such as the type of quality measure used or the size, the incentive or the frequency with which the incentive was given. The implementation factors many on the call I am sure will be familiar with are those CEFR framework categories. Implementation processes are those things that were used to implement the P4P Programs and examples include things like academic detailing, audit and feedback and so forth. The outer setting refers to the broader health system context within which the intervention is implemented. The inner setting refers to the characteristics of the institution or the organization itself and the provider characteristics are just that, demographic characteristics of the providers.

The next category I will just quickly mention, provider cognitive effective and behavioral responses are those things that are thought to be intermediate between the program itself, the implementation factors and then the outcomes. It is the way the programs change provider behavior and their thought processes. So these include things like intrinsic motivation, one’s intrinsic desire to do well for the patient and their cognitive response constructs such as biases; professionalism; identification with the organization and so forth. Then finally process of care and short term patient outcomes are just that.

To jump into the results, there were thirteen studies looking at how different program design features affected outcomes. Overall it seemed that there were some benefits associated with measures that focused on clinical quality in patient experience, while measures that focused on productivity and efficiency were viewed negatively by providers and seemed to contribute to less of a beneficial culture for coordinating care. The second bullet here refers to the fact that the measures themselves, the details of how one constructs a measure of quality a composited measure of quality probably matters. So there are studies looking at fancier statistical models to come up with composited measures were likely more reliable than raw some scores. While it seems like the financial salience of the incentive amount is important it was hard to discern from the literature the magic incentive amount or the percentage of overall salary accounted for by these bonuses that predicted participation or program success.

From our interviews with the key informants, they stressed that organizations should look at a broad range of outcomes that included both process of care and patient outcomes, but that the number of total outcomes mattered. An example of this is that the quality outcomes framework, the QOF study in the U.K. started off in its first inception with a hundred and forty-six measures total and over the course of the next decade they determined that that was too many and they ended up whittling it down to eighty-one. Key informants also felt that measures should be clinically significant, realistically attainable meaning if a measure currently achieves, reaches eighty percent of patients, reaching ninety percent of patients is that really realistically attainable or not; attention to those types of details is important; that measures should reflect institutional priorities; be evidence-based clear and simple. The incentives should be large enough to motivate people but not so large as to encourage gaming in that there is some evidence in particular in the hospital literature that penalties might be more successful then incentives although there was not a great deal of empiric evidence about this. There is probably a need to think more about team-based measures because increasingly there are many staff involved in accomplishing care for a given group of patients. Key informants also felt that timing of incentives should be frequent but balance with the payment size.

The next slide talks about implementation processes. There were eight studies looking at this. In particular one of the questions of interest was – what happens when you continue to increase maximum threshold so in the QOF Program maximum threshold refers to the proportion of patients achieving a certain measure and the percentage that corresponds to the maximum award that a provider can get. Over time they increase the maximum thresholds and this seemed to improve performance especially for those providers who are lower performing to begin with but had less of an impact on providers that had achieved high performance. There was also some interesting questions around what happens after an incentive is removed. There are three studies looking at this and two of them actually were conducted within the VA and looked at, they were respective cohort studies looking at natural experiments where the VA had retired measures or had a set of measures that were actively monitored and then switched to a passively monitored status. They found that once high performance was attained that removing the incentive did not have an impact for their performance. In other words, after incentives were removed performance remained high. The last bullet point highlights one of the studies I just referred to.

Key informants thought that measures should be evaluated roughly on a yearly basis but that they really needed to be considered in an active way in that once incentives had achieved high performance that they should be considered for retirement. The processes used to implement these programs should be transparent and that providers should have easily accessible resources to understand how the measures link to clinical quality and some guidance on how to achieve success. They felt that stakeholder engagement and the bottom up approach getting provider/stakeholder buy-in early on was important.

There were six studies looking at the impact of the outer setting and there were really no clear patterns emerging related to how the region or population density or the patient population affected outcomes. Key informants felt that certainly the needs of the patient population should be considered in design of the programs in that very large multi-site programs should allow for flexibility to adapt to local patient population needs.

The inner setting was examined in eighteen studies. In the U.K. larger practices performed better, but this varied by condition and the indicator used. There were no consistent patterns in the U.S. to determine how the inner setting issues related to practice size, patient volume and so forth related to program effects. In the U.S. it was felt that interventions that focused on culture change in clinical support tools were associated with better performance. Key Informants emphasized that pay for performance should be viewed as just one part of a larger quality improvement program in that other important factors include the infrastructure/infrastructure support, organizational culture, alignment of resources to the pay for performance program and public reporting. Several authors of the studies we referred hypothesized the fact that performance remained high regardless of whether a measure was actively incentivized or not may reflect broader changes to organizational culture rather than the effects of the incentives themselves.

Provider characteristics – five studies looked at that but there were really no consistent evidence that provider characteristics relate to program performance.

Moving on to KQ3, which looks at positive and negative unintended consequences of pay for performance programs. A number of studies, forty-two studies looked at the impact on health disparities and we found no strong consistent evidence that pay for performance programs differentially affected different patient subgroups. They found that groups with lower baseline levels of care tended to improve more over the short term. Key Informants pointed out that in the first two years of the QOF, the U.K. study, there was a reduction in healthcare disparities, but they thought that this was probably due to the fact that there were lower baseline levels of performance in some groups. They look at deprivation areas and once high deprivation practices were high performing, the cost, the incremental costs associated with eliminating the remaining gaps were higher. In other words, initial effects of the pay for performance program at least in the U.K. reduced disparities probably due to very low baseline levels of performance. There were smaller incremental effects the more performance improved. The relationship of programs to health disparities in the U.S. was not very well studied.

Gaming was obviously a concern. There were three studies looking at this but there was no strong evidence, empirical evidence related to gaming. Key Informants on the other hand felt strongly that gaming is likely to occur and that programs when put into place should be designed with this in mind. The last bullet echoes what we have already said that stakeholder involvement, early buy-in, a bottom up approach, evidence-based metrics and realistic measures may reduce the likelihood of gaming.

Risk selection was studied in eight studies. Six were conducted in the U.K. in the QOF and they found that exclusion rates providers were able to exclude patients from metrics if they felt they were not applicable to a given patient. That exclusion rates were for non-white patients, low income patients, and those with comorbid conditions. But that they also found a positive relationship between rates of exception reporting and the total performance. In Taiwan patients who were not enrolled in the program tended to be older and had more comorbid conditions perhaps suggesting some risk selection there. Key Informants felt that in looking at all the QOF data that exception reporting is probably not a huge issue in the U.K. but that there was some concern that algorithms could identify high risk patients and at least the thought then that you could select out those patients. They felt that incentive payments should be risk-adjusted.

Spillover effects - there are both positive and negative spillover effects. There are eleven studies looking at this. The negative side of this is that there is some concern that paying a lot of attention to incentivized measures might take attention away from non-incentivized measures. This was looked at in three studies and there was some evidence of a modest effect here. So again in the QOF and also in the VA the incentivized measures improved over the first several years where as non-incentivized measures by the third year were below pre-QOF trends. On the \_\_\_\_\_ [00:33:24] [audio skipped] evidence to indicate positive spillover effects so this means that there are positive effects that extend beyond the sites that were incentivized and this was found both in the QOF and the U.K. HQID programs. The mechanisms for these spillover effects are unclear but it is possible it is due to provider behavior change or other related interventions such as public reporting or perhaps coincident improvements such as the implementation and improvement of electronic medical records. Key Informants felt that some of the lack of significant differences in one of the studies that Aaron referred to the U.K. HQID program which found that in the long term mortality metrics were no different in intervention comparison groups that perhaps some of the lack of differences was due to positive spillover effects in the comparative groups.

Quickly summing up here - clearly there are a number of examples of pay for performance programs, but as we just briefly touched on here there is an enormous amount of heterogeneity in program characteristics, in the settings in which they are tested, patient populations that are treated and so forth that it precludes us from drawing strong conclusions that can be broadly applied. If we were applying a grade framework strength of evidence rating to our conclusions they should be at best considered low because of all of this heterogeneity. Findings from both the literature and Key Informant interviews suggests that the use of metrics that are evidence based that promote clinical quality especially those that improve clinical quality as viewed by providers and that provider buy-in is a crucial issue. The way to structure incentives is hard to understand from the literature but that certainly the incentive size, the frequency with which it is given and the groups that are targeted for example, positions versus the larger healthcare team need to be considered. Programs should have the capacity to change over time and response time going measurement of data and provider input so the U.K. QOF Program has done a good job of this. They reevaluate their group of metrics yearly and they work in concert with NICE which is their guideline development group and they have made a lot of changes to the program in response to continuous data measurement and provider input. Because some of the data we talked about the fact that removing incentives did not seem to adversely affect performance, that areas which have achieved high performance metrics targeting those areas should be considered for retirement whereas pay for performance programs should really target areas of current or poor performance.

The limitations are of reviewing the literature again. A lot of it has to do with the vast amount of heterogeneity in this literature, the methodology quality of many of the studies themselves and the relative lack of data on hospital based programs, there are some but there is a much larger body of literature in ambulatory studies.

There are a lot of opportunities for future research, a lot of studies have been observational in nature or have lacked good match comparisons groups so there is more need for trials examining these programs. There are very few studies of good quality examining implementation factors and we just touched on some of the factors that might be related to a program success but there is certainly a need for better quality work in this area. There is limited research examining different sub-populations particularly in the U.S., again a lot of the data we looked at was in other countries such as the U.K. so there is a need for more U.S. studies.

With that I will turn it over to Dr. Atkins.

Dr. David Atkins: Thank you Devan. Just a little time check, is Joe Francis on the line yet?

Molly: Not to my knowledge, but I did just send him an email.

Dr. David Atkins: Great. Logically Joe would have gone first to talk about where the VA is going but he has had another commitment so hopefully he will join. I will just begin by reflecting what I think the critical questions for the VA are. I think if you look at the VA, and outside VA, and not inconsistent with the report findings, I think some form of pay for performance is probably here to stay. Certainly Medicare is increasingly tying some of its payment to quality measures. I think there is a philosophical belief that payment should be linked to performance in some way. I think probably although the VA has gotten beaten up for how it uses performance bonuses I think the idea that we would get paid regardless of what quality we deliver probably is a hard sell with Congress.

Then the question is – what direction is the VA likely to go with its current pay for performance and where could research help that. One area is sort of a philosophical question which is – are we paying for performance or should we be paying for improvement. Some people would feel like you are paying for performance and you should not reward somebody who is lousy more because they had more room to improve. Clearly from a systems improvement perspective and the research generally bears it out, you want to try to have the incentives seem achievable for everybody, you do not want people to give up because the target goals looks unattainable.

The second issue relates to the analytic framework that Devan's report used though he mentioned he did not have time to show it. It is understanding how performances, how payment actually achieves better performance. In the analytic framework that they use they note a variety of cognitive factors and process factors and external forces. I think it is important to think that in a broad brush, payment can work in several different ways. It can motivate system changes like putting in electronic health records or changing the actual process of care in a way that does not create extra work that is generally sustainable. That is really the best goal of pay for performance is to motivate people to look at how they can improve their processes in ways that are self-sustaining and equal or more efficient. That would explain why often when measures are removed the performance generally persists.

In other cases we have measures that really are aimed at drawing attention of clinicians to a particular area that it has been hard to get their attention and people have talked about therapeutic inertia. So when we are talking about things like blood pressure control or diabetes control, there is not a simple system solution to that. It is often getting clinicians to overcome the therapeutic inertia and switch a patient to insulin or add an extra medication. Those are measures where one probably is going to see a different response to whether they are there or not. I think what we want to avoid are measures that are all they do is promote extra work or they rely on the extra efforts of clinicians to meet them because that inevitably is not sustainable and it is much more likely to have negative effects by drawing attention away from other equally important issues.

As we think about that framework, I think the VA thanks in large part to Joe’s thoughtfulness is trying to get away from the approach of having too many measures, having measures that are bivariate apply across to everybody without any consideration of individual circumstances or of the things that are the priorities to the site. I think certainly at a national level in terms of measures for network directors, we will see more thoughtful alignment of measures, where measures that clinicians are facing, I think there are a number of areas of research. One area which we already are funding at a fair amount of work and have gotten some important results is just getting better measures and that is really a separate question from pay for performance. But some of the problems that Devan’s report eluded to in terms of negative effects, are not necessarily negative effects of paying for performance they are negative effects of paying for performance on the wrong measure. If you have a measure that encourages a clinician to lower the hemoglobin A1C to a target without taking into accounts the individual patients circumstances, that is going to have unintended effects, not necessarily the problem of pay for performance it is the problem of not having measures that are individualized. People like Eve Curran and others have done a lot of work in thinking about how we get measures that are more patient centric and measures that do not promote over treatment.

The second issue, which is a little more specific to pay for performance, is how do we choose the measures. And I think the report made recommendations that I think research should try to confirm which are the value of rotating measures of matching measures to areas of low performance rather than having the same set of five chronic disease measures across all sites but allowing individual sites to choose the measures that they want with an improvement mindset rather than an accountability mindset. To see whether that works better than just applying the set uniform set of measures across the system.

The third area where although I confess, I do not know how much flexibility the VA has is the issue of payment structure. We talked about the report alluded to the fact that people value losses more than gains and so penalties are more effective and that certainly is the way Medicare is approaching things in the hospitals where you can get panelized for readmission rates. That is not likely to be feasible on an individual scale and it is going to be challenging in the context of the VA to how much they will let us play with different payment models to commission’s. But one could imagine offering different types of reward structures where the total pool is the same across different groups of commissions but it is targeted in different ways and testing it in that.

Then the last area of research which would probably be even more challenging in the VA but people like David Ash, Kevin Bolt have done interesting work on is the role of the patients. Certainly, in chronic disease there is a patient who is as much in control as the clinician in terms of treatment decisions and adherence. The groups at Penn and Philadelphia VA have done interesting work on the combination of patient and provider incentives. If my memory is correct the combination of both works better than either one if you are looking at something like hitting with the target.

I will stop there and see if Joe is on the line and if not we can open it up for questions.

Molly: Thank you Dr. Atkins. I am in communication with Joe and he should be joining in just about ten minutes so he will probably be giving our wrap-up comments. Before we do move on to Q&A I just want to check with Karli or Cheryl to see if either of them would like to provide some comments before we move on to Q&A.

Cheryl Damberg: Excuse me, this is Cheryl Damberg. I just wanted to congratulate this team on doing a really terrific job summarizing what is a very diverse and challenging landscape of articles to make sense of. As was noted several times in terms of the work that has been done to study pay for performance and natural stetting’s, have revealed a lot of diversity and its heterogeneity is not well captured in terms of the description of the studies. I would underscore the point of future research really needs to do a much more systematic job of capturing the design and contextual variables for any given pay for performance intervention such that both of us were trying to learn from any type of experiment that is in play can really understand a lot of those factors that may be determinants or mediators in the success of an intervention. I would definitely highlight that.

I think the other thing to note is as I studied this landscape, there are a number of key dimensions so one has been touched on already which are measures and when we have sort of the right types of measures that really are tailored for individual patient circumstances and I think that is right on. I did not hear comment about what I call the potential for mis-measurement. I think that there are opportunities to try to look at adjusting for differences in patient mix across different providers to try to make those comparisons more fair. I think in doing so that can help guard against one type of unintended consequence which is avoiding patients who are going to be more challenging, if you think you are not going to hit the target that the program specifies.

The second thing and it is related to this point I was just making is that the payment structure is often overlooked in terms of how to better incentivize providers. Typical pay for performance programs has paid either for attainment of some target or based on a relative basis or have paid for attainment and improvement. However, in recent work that I did with a colleague here at RAND, Morgan [ph] Elliott. We looked at the extent to which providers who care for disadvantaged patient populations and who may be under resourced from the gecko absent any type of pay for performance, they would be further disadvantaged by these programs absent taking into consideration some of these factors. That article was published in *Health Affairs* in January of this year and I encourage people to think about how we consider the payment structure such as it does not lead to undesired consequences.

Then the other thing that I would note is that, and I think why this work is so important, is that there are a lot of what I am going to call structural or system characteristics that help shape the design of these programs. And that landscape is not well understood as well as those system and structural support mechanisms can help a program like pay for performance targeting selected measures be more successful. Because they potentially offload some of the burdens from the frontline physician to assure that these actions happen. Now as was mentioned earlier on today’s call, not everything can be systematized. But I think that there are many things that when I looked at the measures contained in pay for performance that organizations, the health systems in which these providers work are trying to own it at the front end to make sure that those prophecies have not, independent of what happens at the frontline of healthcare.

I am just going to stop there, Karli.

Dr. Karli Kondo: I actually do not have anything to add, thank you Devan for stepping in for me I very much appreciate it.

Molly: Wonderful, thank you to you both. So we are ready for Q&A now, for our attendees that joined us after the top of the hour to submit your question or comment for any of our panelists please use the question section of the Go To Webinar Dashboard, just click the plus sign next to the word questions and that will expand that box. We will get to those in the order that they are received. For the first question actually Devan do you want to advance to the last slide?

Dr. Devan Kansagara: Sure.

Molly: Okay thank you.

Dr. Devan Kansagara: There you go.

Molly: Perfect. For the first question – what are your thoughts on the notion that if Veterans had maximum choice to choose their healthcare provider, VA or private sector, then pay for performance would be a moot point because the Veterans would be the manager of the contingency of pay for performance?

Dr. Devan Kansagara: This is Devan. I would be curious to hear others thoughts as well. It is a really interesting question. It gets a little bit into I think the notion of the comparative effects of public reporting and pay for performance. So essentially I think the question implies if you could really use public reporting in a way that is instrumental in the VA and currently it would not be because if you publicly reported and they do not have a choice in where to go, then there is no way it can have an effect. We did not look specifically; our review did not look specifically at public reporting as an intervention. I know there are some studies suggesting some promise in the notion of public reporting. Actually, it was interesting in the Key Informant interviews, correct me if I am wrong Karli, but there were a number of Key Informants that brought up this notion that public reporting may be an important factor that explains secular trends in healthcare that might have washed out some of the pay for performance effect.

Dr. Karli Kondo: Yes, that is absolutely true, there are a few Key Informants that very strongly stressed that they thought that public reporting was or may be more powerful than the actual incentive. Some of the case studies one of the things that were noted was that even prior to the QOF and prior to some of the hospital demonstrations that there was improvement in both the hospitals that were involved in the studies as well as the overall potentially just because it was known that the pay for performance was coming.

Dr. Damberg: This is Cheryl to add on to that because we did some of the evaluations of the early hospital pay for reporting. We did interviews with hospitals and they were very concerned about their reputational effects of these programs. If you look at when hospital value based purchasing went into effect it was long after the public reporting aspect most of the gains had already accrued. And hospitals that were not in the hospital quality incentive program which was a pilot test of pay for performance the format of care went live with that, the hospitals were essentially shadowing the participants if you will, that kind of spillover effect, anticipating that there was going to be pay for performance more broadly. I think that this public reporting is a very powerful incentive.

Molly: Thank you for those replies. The person who wrote that in this was during Devan’s answer said – yes if that is way of quality indicators then their actions in the free market of healthcare would serve as pay for performance. Thank you to that submitter. Another person writes – I joined late I apologize if I missed this in the beginning but will this report be published publicly?

Dr. Devan Kansagara: Yes, the report is currently available on the VA intranet and it will be available publicly November twentieth I believe. The first few slides have some links to the VA ESP site but you can also just Google VA ESP and the first link is the ESP program site, which includes all the published reports. In addition, the group is in the process, there are two manuscripts in different stages of production that will hopefully be published related to different aspects of this report.

Molly: Thank you for that reply. Here is kind of a loaded question – do you see any policy changes being implemented straight away based on these report findings?

Dr. Devan Kansagara: This seems like a good segue if Dr. Francis is on the phone now because I think that is part of what he was going to touch on.

Molly: He should be joining us any second he is with the Undersecretary for Health so he should be here just momentarily. We do have another, go ahead.

Dr. Devan Kansagara: Could we come back to that question then because I think that is kind of exactly, what he was going to talk about.

Molly: Absolutely. The next question we have – I have noticed with a lot of EPS, I am sorry ESP reports, they come to the conclusion that more research needs to be done. Here is Joe we will let him call in give him just a second but we will finish up this question. Dr. Francis are you on the call?

Dr. Joe Francis: Yes, can you hear me?

Molly: We can thank you, you may go ahead, I am kind of throwing you on the spot, but we did just have a question about if any policy will come from this report but we also welcome any general comments that you would like to give.

Dr. Joe Francis: Sure and apologies. Actually, I was in a long meeting with some of the senior leadership of VHA today probably a direct result of some of the failures of our performance measurement system. Namely the access crisis which in many ways was a classic illustration of the problem with performance measurement which is the human tendency to mistake means for ends like we were meeting the needs of Veterans by figuring out some way to make this artificial fourteen day waiting time goal. That was actually the genesis of this request to the ESP program. In other words, what do you know about how one should implement performance measurement and pay for performance systems. I was not here for the earlier conversation but I think it is fair to say that you probably learned that we do not know a lot about how to do this right. We have some ideas from the Key Informants, the people that were involved in the studies, but those make some fantastic questions and hypotheses for future research.

Because in the end I think performance measurement and performance management, is less about getting the measure and getting the data right and much more about how you get the motivation right and involve it properly and use all those principles of social science to align people with what the organization needs for success.

From the policy standpoint, I would say a few things. Number one – it is our goal to be smarter in policy making and in driving policy and that is very different than having more policies. We tended to have a lot of policies just like we have had a lot of measures and there are often reasons for that in a system where boxes need to be checked for various accountability holders. We have already taken some of the insights from this work to stand up a new group called The Performance Accountability Work Group that is really thinking quite thoughtfully about what our measurement system should look like for VHA. I can tell you just a few general principles of that - number one is you should not have too many measures because it creates confusion and a diffusion of efforts.

The second is that measures can be applied or should be applied differently at different levels of the organization. Take the example of the access; it is absurd to apply to a scheduling clerk the performance standard of fourteen day access for all patients. Because number one that scheduling system has no way of achieving that goal if there are no provider slot availabilities and the goal itself is rather arbitrary it makes no sense. I would rather have a performance goal for a scheduling assistant something along the line of - you will pick up your phones within five rings and you will ensure that messages get turned around by the appropriate person within twenty-four hours. Because we see in our surveys that is like a really, really big deal, people are not so upset about not getting appointments in the VA as they are about not getting phones answered or phones returned. That is at least one principle that we think right measure to the right unit of the organization.

The second is really about engagement involvement with the people that are affected. A lot of our measures are basically brought down from Mount Zion, carved in granite. In fact, what we are trying to do now is to reach out to frontline leadership engaged in this process of helping us go through our existing measures and decide how those things should be applied. That itself is not easy to do because a lot of things happen in our organization and all of those discussions take a lot of time. We are not expecting to see the result of those efforts in place any time soon but we are aiming to have at least the 2017 performance plan reflect a more evidence based and thoughtful approach.

The third policy level and this is an issue that we see in Washington where every measure, every policy has its own constituency groups. We are saying no to people in other words. We know from your perspective of the universe that you are the center but in the global perspective right now, the things that threaten our organization are relatively few in number. The Undersecretary is in fact articulate those in the last week so access is key. Employee engagement is key. Our ability to create a high performing network of providers in the community is really key to our success. The ability to identify and disseminate best practices is really, really key. Addressing things like our leadership gaps so that we have the right people running our healthcare organizations, those are really the five areas that are kind of the short list that we have to focus on and that is not negating the importance of all the other measures that we know and love and do not necessarily disappear. If we do not get these core things right we are not going to be around in a few years to get the other things done.

Molly: Thank you very much Dr. Francis we appreciate you providing your comments.

Dr. Joe Francis: By the way, I have one more thing, one other priority, which I actually forgot about and how, could I? In the end, it all comes down how the trust the Veterans put in VA care and no one is saying they do not trust us because they do not like how we look on performance measures. They feel they do not trust us because we are not very consistent in honoring our promises the things we say we are going to do, we need to do time and time again very, very well. It is better to have a shorter list and do those consistently well then to keep enlarging the list and adding on to almost impossibly huge plates for our staff to consume.

Molly: Very well put and I am sure any VA employee would back you on those statements. I do want to give the other presenters an opportunity to chime in on the comments that Joe made or any concluding comments. David did you have anything you wanted to add before we move on to the report presenters?

Dr. David Atkins: No.

Molly: Alright. Devan do you have any concluding comments you would like to make about the report or just anything in general before we wrap up?

Dr. Devan Kansagara: Just quickly, there was the other question that had come in and I did not want to ignore the other question. There was a question about I think it was getting at that a lot of our reports end with more work is needed which is true and just quickly addressing why that may be the case. In part….

Molly: If I may interject, I am sorry, but can I repeat the end of the question.

Dr. Devan Kansagara: Yes.

Molly: They just wanted to know if you do follow up reports based on these topics that have been introduced.

Dr. Devan Kansagara: Yeah, yeah. No, it is a great question. That is an issue \_\_\_\_\_ [01:08:13] [audio skip] in general for evidence review work, when do you do updated reports, what is the timing of that and we do not have a firm system in place for that. After several years have elapsed if it is known that a body of evidence has accrued there may be opportunities to update reports. There are certain groups, systematic review groups that do have a process in place for doing that and it may be something that we end up doing in the future. Part of the reason for doing things like this seminar and having folks like Dr. Atkins and Dr. Francis join us is that there is more of an opportunity for us to just present the evidence and not stray into providing recommendations beyond what the evidence shows. Then also allow for some practical discussion about how do you tease out some actual implications from this work. The only other thing I will say in summary is to again acknowledge all of Karli’s hard work on this report and her information is there if you have other questions and thank you.

Molly: Thank you so much. Karli did you want to say anything before we get going?

Dr. Karli Kondo: No thank you very much for everyone.

Unidentified Male: I do have a question. Do we know when this report or some version of it will appear in the peer review literatures so that we can communicate this to broader audiences?

[01:10:03] [many speaking]

Dr. Devan Kansagara: Go ahead Karli.

Dr. Karli Kondo: I was just going to say that we have two manuscripts that are currently in process. One is currently under review and we are in the final stages of drafting a manuscript. The one that is in review at the moment focuses on the implementation processes and we are in the process of drafting a manuscript that looks at Key Question One the effectiveness question.

Dr. Devan Kansagara: Outstanding.

Molly: Thank you. Not to be forgotten Cheryl or Aaron do you have anything you want to add real quick? Okay. I just want to thank you all so very much for coming on and lending your expertise to the field. I know we tried to fit a lot in in our short amount of time but I think you all did an excellent job. And of course I want to thank our attendees for joining us, we really appreciate you spending this time with us and I am going to close out the meeting in just a second a feedback survey will populate on your screen, please take just a moment to fill out those questions. We do look very closely at your responses and it helps us to improve sessions we have already provided as well as give us topics for new sessions to facilitate.

Once again thank you to everyone and in case you did not see my poll slide plug please check out VA polls and join us so we can continue the research conversations there. Have a great day everyone, thank you so much.