NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
August 15, 2012
2:00 p.m. ET

Operator: Welcome to the conference. Please note today's call is being recorded. Please standby.

Reva Winkler: Good afternoon, everybody. This is Reva Winkler from NQF. Thank you all for joining us on today's conference call of the Infectious Disease Steering Committee Workgroup.

Today we are going to discuss seven of the measures in this endorsement maintenance project. The purpose of this call is for the members of this workgroup to share their preliminary evaluation and thoughts about the seven measures we have for this workgroup.

We, this morning – yesterday and this morning, we sent updated version of a summary of your preliminary evaluations, those that you had submitted via the survey monkey tool. So, hopefully you have that. That's also what we're going to be displaying on the webinar. This might be a good point of – a beginning point of discussion as we look to see where people agreed or potentially disagreed or where there might be questions.

So, so far, I have other workgroup members, Adam Thompson is with us, Michael Farber is with us, Ed Septimus is with us, and Kalpana Ramiah is with us. Did Peter Havens join us yet? OK, we're still waiting for Dr. Havens.

Ed Septimus: Are there developers on the call?

Reva Winkler: Yes, and I was just going to say, we have the folks from NCQA and also for HRSA. Now, this is a public call and so, there may be members of the public listening in as well. At the end of the call, we'll open the lines and give them any opportunity for comments should they desire.

And as the operator noted, we do record this call because frankly, we find the transcript very useful writing up our summaries.

Michael Farber: (Adele), this is Dr. Farber. I'm having trouble, you know, the e-mail that you sent today which has at the bottom one of the attachments, the log on site won't let me do it. I can't enter my name.

Reva Winkler: The webinar address you mean?

Michael Farber: Well, you sent an e-mail today and the e-mail has a link, you know, it gave the number to this conference call and then the link to the webinar, but it won't let me put my name on it. It won't respond to it.

Reva Winkler: OK. So, we're thinking it might be something with your organization's computer setting.

Michael Farber: Is there another way that I can log on and what should I be logging on to?

Ed Septimus: You might – this is Ed. What you may want to do is try to copy and paste the URL into your web browser as opposed to directly from the link. Do you understand what I'm saying?

Michael Farber: I have some help here. Do you know what that is?

Female: Yes, I could try that. It's just that when we get to the log in section where it says Enter Meeting, that's one we're trying to enter in information into those fields, like Dr. Farber's name and all of the other information in those boxes to enter. But once – every time we try to type into that, something kicks us out from that field.

Ed Septimus: OK, that's a little different.

Reva Winkler: Yes, it is. I'm not sure what we can do help you as yet but what we're showing on the webinar is the document, the Word document that (Adele) has sent this morning. Do you have that?

Michael Farber: Yes, printed.

Reva Winkler: Then, that's what's all that's going to be shown. You're not missing anything else.

Michael Farber: OK.

Reva Winkler: OK.

Kalpana Ramiah: Hey, everybody. This is Kalpana. The last e-mail I got from (Adele) was last night at 7:00. Is that the same document that was sent this morning?

Reva Winkler: Yes, that's the document we're working off of.

Kalpana Ramiah: OK.

Reva Winkler: All right, folks. So, has Dr. Havens joined us? OK. If not, well, maybe perhaps move the first measure on our list to a little bit later and (inaudible).

The first four measures we're going to talk about today all addresses similar topics and that's around medical visit or (inaudible) visit. We have one measure that's previously endorsed from NCQA and then there are three new measures submitted by HRSA around this similar topic. So, they each approached it – the subject in a different way.

So, what I propose we do is discuss each measure individually and goes through the evaluation criteria beginning with importance and then moving on to the others. And so, since Dr. Havens isn't here, Adam, if you would …

Peter Havens: I'm here now, but …

Reva Winkler: Great.

Peter Havens: I'm walking from the hospital. So, if you want to start with Adam that'll great.

Reva Winkler: OK. Super. Yes, we'll do that. We'll just move yours to the fourth one, I think. Adam, if that's OK with you, let's start with measure 2079. This is a new submission from HRSA and this is a measure of medical visit frequency. The level of analysis for this is the clinician group practice or facility, and essentially it's the percentage of patients regardless of age with a diagnosis of HIV who had at least one medical visit in each six-month period of the 24-month measurement period with a minimum of 60 days between medical visits.

So, if we look at the importance criteria, it looks like various members of the workgroup had various things to say. Adam, would you like to share your thoughts on, say, impact and the opportunity for improvement on performance gap to start with?

Adam Thompson: Yes. So, when looking at – let me pull up my paperwork so I make sure I'm looking at the right one.

We're looking at the importance to measure, there was significant data presented as far as looking at the importance of medical visits over a period of time. They presented information showing that there was a decline in individuals making their medical visit as time passed on. They showed up at a six-month, 12-month, two-year, and then three to five-year interval.

When they were citing their performance gap, they showed that there was significant room for improvement, that overall under 50 percent of the patients were meeting the HRSA criteria and for medical visit frequency. They do have their data stratified by disparities and they identified that, and the importance to measure which will come up a little bit later.

So, in the importance to measure, when we get down to looking at the evidence on this, they had quality of studies was listed out saying that the systematic literature review was evidence based restricted to randomize controlled trials and observational studies. And they specifically identified that the recommendation that they're basing this on was focused on monitoring retention to care and that was based on two studies.

They then go on to the cite that the Department of Health and Human Services, 14 studies and then eight studies examining impact of treatment on prevention of transmission. So, it was not just something that was of importance to the patients but also for public health priority around ending the spread of HIV.

When it got to the quality of body of evidence, it was described as two well-designed analysis of cohort studies. Throughout the presentation of this, I was looking for more than a single sentence, so I was the one that actually rated that the quality insufficient. But their consistency, they cited the two studies that they were comparing, what the difference was and what the confidence interval was and both of those studies were consistent.

They cited that the only real cause to this was going to be around the attending of the attending of the medical visits and the costs associated with that medical visit but that there was no perceived harm in a person attending their medical visit.

They went on to describe the body of evidence has been graded by the International Association of Physicians in AIDS Care. They used to modify grade system which was one of the NQF's recommended grading systems. The grades were assigned A1to A3 which indicates that it was a good body of evidence.

They go on to quote the guidelines around this and in this measures, what was interesting about their evidence is they specifically talk about retention to care and the frequency of visit over time versus some of the other measures that we'll talk about later which was specifically more to the patient making their visit, but this was looking at how long over a 20-month period or 24-month period were they consistently making their visits.

They did identify that this was in response to the fact of the guidelines they chose because there were no other guidelines addressing toward attention to care. The measure steward themselves rated the quantity as moderate, the quality as moderate, and the consistency high.

It looks like most of us agreed with that assessment. Some of us rated the quantity as high. The majority of us rated the quality as moderate, and all of us thought the consistency was either high or moderate.

Reva Winkler: Super, thanks. Any other thoughts or comments from other members of the workgroup? Everybody thinks that's a fair summary? Do you feel that it looks like there was only one person who really felt the evidence wasn't there, does that person maybe want to comment?

Ed Septimus: This is Ed. I have one question about exclusion on this. Let me see if I can find it. Can I ask you where you're seeing everybody's evaluations from this group?

Reva Winkler: This is on the summary documents that were both showing on the webinar as well as we emailed it to you last night and this morning.

Ed Septimus: Under the denominator exclusions which is page 9 of the document you sent out, the exclusion is for the patient who dies anytime during the 20-month measurement period, so as with the other documents as you know we may talk about harmonization of some of these, talks about incarceration or other reasons for not coming back to visit. Should that be included as a denominator exclusion?

Reva Winkler: Would you like to ask the developers to comment?

Ed Septimus: I'd like to ask the developers to comment because you see it in some of those measures but not this one.

Marlene Matosky: So, hi. This is Marlene Matosky. I'm from the Health Resources and Services Administration HIV/AIDS Bureau. I'm on the team that participated in developing this measure. Originally, when we had done some additional work with this measure in the field, we had exclusions related to incarceration, relocated, or transferred service.

And as we had some of our providers utilizing these measures, we received feedback that it was difficult to provide structured data related to those three elements. And in addition, Larry – sorry, Kevin Larsen who is the medical

director from the Office of the National Coordinator for Health I.T. has been saying that you know, both his office and CMS are saying that we need to (inaudible) more measures so that we're able to specify all aspects of the measures for use in health records. So, that's the reason why we eliminated those particular exclusions from these measures.

Ed Septimus: One other question, I think this is one that's difficult. What do you do about non-complying patient for instance, if someone missed their appointment twice in a row? There are some – the U.K. uses this actually. It's (inaudible) non-compliance excluded from analysis.

Marlene Matosky: I guess I would respond to that in two different ways that both get to the – to your questions. The first is – and we haven't indicated any sort of benchmarks or goals for these measures, but certainly, we don't ever expect anything to be zero percent or 100 percent. So, that when we do start setting benchmarks – and CMS has really spoke to this eloquently through the medical or through the meaningful use project, is that once we start getting in and we start collecting national data on measures, then we'll set benchmarks but we will take into account these sorts of items that you're bringing up as potential reasons why somebody might not come in for visits.

And then in addition specifically to you know, the idea that folks might be you know, not in compliance and the physician can't reach out to them to bring them back in for a visit, you know, we feel like, you know, part of quality – you know, the flipside of performance measurement is that quality improvement piece. And so that, you know, we're hopeful that instead of just using these as measures, folks will do that quality improvement piece and make attempts to, you know, bring folks back into care.

Ed Septimus: I think that's an excellent point because when we talk unintended consequences as we see in other core measures, people try to gain in the system and try to make it look perfect. This is the first time I've heard a medical person saying we don't expect 100 percent. And we may be looking more in improvement than actual 100 percent compliance. But often, it's viewed differently.

Marlene Matosky: You know, sir – and you know, it's not just here at the HIV/AIDS Bureau that feel that way, it's also our colleagues over at CMS and ONC are saying the exact same thing around setting reasonable – you know, I'm not saying like easy to step over, you know, the thresholds and benchmarks, you know, something that you have to stretch to attain or work you know, to attain.

But this isn't just about performance measurement for us. There's also that other side of quality management which is the quality improvement piece of it. So, we very much accept this as you know, a two-way, you know, process; the measurement and the improvement.

Ed Septimus: All right. Is there some way that that can be highlighted? I mean, I completely agree with what you just said. But that's not always how these measures are interpreted.

Marlene Matosky: So, I guess I would go back to just ask our leaders over at NQF, is there something that we would put in the end of the document in the additional comment section?

Reva Winkler: Well, or in your rationale for the measures. There are many number of places you could, you know, insert that. But, yes.

Marlene Matosky: OK. Thank you very much.

Ed Septimus: I think that would be helpful. I'd like to hear other people on the call as to whether they think that's a reasonable thing to include.

Adam Thompson: This is Adam. I would agree, yes. I would definitely letting people know what the expectation is as far as performance and that they're not going to be, you know, measures necessarily on non-compliant patients, really and I think gives providers the viewpoint that they can feel more comfortable not only with the scores that they get but also if they have an unblinded data system sharing those scores with others.

Peter Havens: And, this is Peter Havens. I would take exactly the opposite approach, sorry. But the reason this exists is to identify – to help programs identify that they may have a high proportion of non-compliant patients and stimulate them to

get those non-compliant patients back in the care in ways that may be a bit of burden and more expensive on the healthcare system trying to deal with patients like these.

So, from my perspective, these performance measures exist so that, right, we can learn what's happening in our practices and if we see that instead of having 90 percent followup over a two-year period, we have 70 percent followup over a two-year period and these measures should make our clinic change dramatically so that instead of saying these patients are all non-compliant, we change our clinic so that it will get patients into care and keep them into care.

So, I would argue really the opposite. Now, what (inaudible) and other evaluators do I agree is abysmal way to get negative and accusatory but from the perspective of a clinic manager, we have to use these data to say, "We're not doing very well and our patients are not staying in care. We need to do something different."

Ed Septimus: I'm not sure that we're setting anything different, actually. I think this is more of a …

Male: I agree entirely with the last speaker. I think that to take out the non-compliant patients would give very high numbers of success which would not be realistic, and I think that it's an important benchmark to be measure from year-to-year. So, I think that – I agree that it's an important thing to include.

Ed Septimus: I think you misunderstood my comment. I mean …

Adam Thompson: I thought like the last speaker, I thought that you were suggesting that if there's a non-compliant patient who's not coming back that they would just be excluded from measurement. I misunderstood.

Ed Septimus: No, no, no. I discussed that some people do exclude that but I think trying to look at this as P.I. tool, I get – what my comment was that I don't think people should look at these measures that you have to comply a 100 percent. But the purpose of it is not only to have obviously good compliance but it also acts as a P.I. tool.

Adam Thompson: You bet.

Ed Septimus: So, I was not – I was just speaking what the U.K. does but I'm actually agreeing with the last two speakers, you included, that you make it clear to people looking at these measures that we're not expecting a 100 percent.

Adam Thompson: I misunderstood. Sorry. I absolutely support that.

Reva Winkler: OK. Do we move on this measure in scientific acceptability, reliability, and validity? Adam, do you want to share with us your thoughts?

Adam Thompson: Sure. So, as far as reliability and validity goes, based on the NQF endorsement criteria, because it was only tested that the measure score, the highest rating that it could get on either was moderate. Because it wasn't EHR tested, they weren't required to show reliability. They could've submitted validity but they did both. And the description of the numerator statements, they go in to describe it as a 24-month measurement with the numerator window being the 24-month time.

And the denominator statement, one of the things I thought would go that's required that they document the test, so that it can't just be assumed. So that it actually have the date that it occurred. Also, when looking at this, the data sample that they presented for their reliability testing was representative of HIV incidents nationwide based on the CDC data from 2009.

They also indicated that the sites were representative of both academic and community-based HIV care as well as the geographic divisions of the United States, as well as demographic diversity of the infection.

Also included was insurance coverage and status which when you look at their data, it shows how important it was to sort of add that. The analytic method they used was looking at a signal-to-noise ratio. They presented the data across the site that they were sampling. Three of the sites were kicked out because they weren't submitting regularly over the time period they were analyzing and I believe, if I remember correctly, the pediatrics sites were joined together because of the low number of patients.

When you look at the validity testing, they measure sort of how they were doing it matched up with their importance to measure evidence. Again, they presented the same data sample showed sort of across over three years with the patient number increasing over the year.

The analytic method that they used was to establish face validity. The technical work group used to modify Delphi process which was one of the NQF's recommended processes. And the workgroup members (inaudible) important, usable, and feasible and this was – they reviewed seven measures and this was one of the four that actually came out.

They also had a webinar and feedback that they got the written responses from (inaudible) providers of the persons who are actually using this. They go on to identify that the testing was not performed on excluded patients. So, when you look at the threats to validity, there really wasn't any discussion of that.

They present their data sample again and as was asked in the result, they show what were the quartile mean and medians of the data results. They've (inaudible) to show their disparities in care data which showed that there were disparities in care across the site as well as across some of the demographics particularly around uninsured.

So, because of that it looks like most of us rated it moderate across the board with the exception of one of us who scored it as high. But unless I was missing something because it wasn't tested on both of those data sources, I'm not sure if we could give it a high score on reliability.

Reva Winkler: Correct. Any other thoughts from folks? Do everybody generally feel comfortable that this measure has been demonstrated to be – produce reliable and valid results?

Any other thoughts from other workgroup members? I'm going to assume silence as agreement. Then in the interest of time, Adam, what about the criteria of usability and feasibility?

Adam Thompson: All right, so as far as usability is concerned, the technical workgroup that did face validity also saw that there was utility and publicly reporting this. It's currently being used in the National Quality Improvement Projects focused on retention and medical care. And upon potential endorsement (inaudible) as well as physician quality (inaudible).

There was a use of the data, it looks like across several major government agencies as well as private healthcare providers, as well as the 12 cities project, and they identified that this measure has been put forward to fill the measurement gap for retention of HIV care.

They showed that there was a meaningful gap and that it was useful. They went on to again identify the same quality improvement project as well as that this measure was unique and that it responds to recent literature regarding retention and risk (inaudible) to health outcomes.

When looking at feasibility, the data was generated during the care process. Everything is available on electronic sources. They identify to their knowledge, there were no known inaccuracies or errors, and as far as their data collection strategy, this is where the discussion came in as far the exclusion of incarcerated and transferred to care.

And they identified that they were pulling the data again from the 15 sites that they had used for the reliability and validity testing. The only concern that I had raised around the susceptibility to inaccuracies would be the comparison of what's in the electronic health record to what was actually in the chart. But there was no real discussion of that.

When looking at our scores, two of us rated usability as high, one in moderate. And feasibility was the same breakdown of two as high and one moderate.

Reva Winkler: Thoughts from anybody else on the workgroup? OK. So, I think – does anybody have any questions of the developer about this measure or about the criteria or anybody of your colleagues comment before we move on to the next measure? OK.

Ed Septimus: Adam, this is Ed. I think that was an excellent summary. Thank you.

Adam Thompson: Thank you.

Reva Winkler: All right, let's move on to measure 2080 which is gap in medical visits. Also, a new measure submitted by HRSA and very much related to the ones we just talked about the percentage of patients regardless of age with the diagnosis of HIV in the last six months (inaudible) year.

Reva Winkler: Just so everyone knows, we're actually streaming the audio from the webinar. So, if your computer speaker's on, we might have an echo.

Reva Winkler: Yes, OK, good. OK. So, this is Dr. Farber's measure, correct?

Michael Farber: Yes.

Reva Winkler: OK. Would you like to start out with the impact and opportunity for improvement, and evidence discussion?

Michael Farber: Well, the impact was felt hard by everybody, and I – and I agree with that. The second issue was …

Reva Winkler: Opportunity for improvement, performance gap.

Michael Farber: Everybody said that was high or a big opportunity for performance gap. Obviously, this is something that are easily measured but need to be encouraged.

Reva Winkler: OK.

Michael Farber: So, the numbers actually on most of the studies were in a range that was in the 60 percent, 50 percent. And so, I did think that there was certainly a tremendous opportunity for improvement in this measure.

Reva Winkler: OK. And then the evidence, is the evidence cited different than the evidence we talked about on the previous measure?

Michael Farber: The evidence is very strong. A lot of studies has been done of course on visits, and the evidence of – their value is very high as far as patients

following care, mortality, getting early treatment for antiretroviral therapy. So I think that that's a very high area.

Reva Winkler: Thoughts from anybody else from the workgroup? Any other comments?

Peter Havens: Hi. It's Peter Havens. Are we asked to try to comment on how similar this is to the one prior, how this one might or might not, and the more or less valuable than the one we had just previously discussed?

Reva Winkler: Ultimately, definitely, you're going to do that at the in-person meeting. Now, we really want to go through them individually to determine if there are any particular problems with any individual measure that would cause it perhaps to you know, dropout of your recommendations.

But because they're so similar, I think they're you know, it's not unreasonable for you to begin to start thinking about how they look at the date similarly or differently, and how that relates to the evidence as well as the value of the information that's obtained.

Peter Havens: And so, the question I guess I would ask is as we review the quality, the evidence that it goes along with looking at visit frequency, are there comparative studies that suggests a relative value of being seen (inaudible) this year compared to more frequently over the next two years. Do you understand why I'm saying?

I'm trying to sort of identify what might be the essential difference between 2080 and the one just prior. This one has a 12-month time period and if you've been initially, were you seen six months later, as opposed to if you were seen initially, were you seen regularly over the next 24 months which is the prior one.

So, the question I would ask from the perspective of the background information is are there date to identify that a two-year time window is a better indicator of an adequate process of care compared to a one-year time window. And does that matter to us in the long run, I guess?

I'd be glad to have the question criticized because I'm not even sure that's the kind of thing I'm supposed to be asking.

Reva Winkler: That's a perfectly fine question. I think what you're asking is the different approaches the two measures take particularly around timeframe and how the data is captured, how does that relate to the evidence that supports the underlying concept of retention that is related to patient outcomes.

Peter Havens: The evidence reports that the retention in care is useful – is linked – the retention in care is linked to patient outcome. That's what I'm going to say that the evidence suggests. Now, the questions becomes is the evidence robust enough to allow us to say that retention in care measured over a two-year timeframe is more closely linked with long-term patient outcome than retention in care over a one-year timeframe which is overstated but sort of the central difference between these two measures?

Reva Winkler: Thought from any other workgroup members?

Michael Farber: This is Michael Farber. I agree with the problem with that. I had the same issue when looking at all the measures related to visits. And that is that we're being asked to really look at this one measure alone, but even in this measure, you know, a lot of time, measures take the minimum. In other words, the four months is the minimum. There wasn't a lot of ability to see studies of comparing different time periods, in other words with two months – every two months gives you better or every three months because every three months relates to getting antiretroviral – I mean, viral load studies and CD4 counts.

So, I think that that's really – I found that as an issue too and that we had several measures that are kind of getting at the same thing and that's retention, and yet they're all a little bit different and they all show a benefit because I mean it's – I think it's also somewhat illogical to believe that visits are going to be useful for people who have complex diseases that need counseling, monitoring, and discussion.

But what we don't know from this is what is the optimal range of visits that give the best data?

Adam Thompson: This is Adam. I had a question around this when comparing all of them specifically the three HRSA (inaudible) measures to the currently endorsed measure was around the definition of a medical visit. And the one that we'll discuss later they currently endorsed when they specified that a medical visit with specific individuals.

The HRSA measures don't give a definition of medical visit and so this may be a question to the developers but what was – what were they envisioning as far as a medical visit? Could it be a visit with a clinical social worker who's setting up care, would that count as a hit in their system? Because then I think when I was comparing them, I compared the gap in medical visits to the currently endorsed measure whereas the other two measures I sort of pulled them out because of the frequency window because I think they were looking at something a little different. That's how I compared it.

Reva Winkler: Comment from HRSA?

Marlene Matosky: So I think I heard two – two sort of baskets of comments here. I think the first one's related to – we had put forward three different measures to – three different measures that's related to retention and we looked at retention from you know, a couple of different angles.

And much of this was driven by some recent work that was done by (Mike Mogavero) where he looked at a variety of different retention measures and looked at you know, the pros and cons that (inaudible) to those measures and you know, he came to the conclusion in his paper that you know, retention is one of those concepts that you may have to look at a couple of different ways to really the most information about your, you know, patient population.

Then specifically towards you know, is there and I'm going to try to kind of paraphrase what I heard Dr. Haven say followed up by Dr. Farber, is there you know, kind of a particular timeframe whether it'd be a 12-month timeframe, no visit in the last six months, two visits in a year, four visits in 24 months; is there something magical about anyone of those in particular?

And I would say, you know, (inaudible) you know, wrote an article where he talked about you know, an increase in baseline CD4 count, you know, with

significantly higher with optimal retention over 24 months as opposed to folks who didn't have optimal retention.

And when we're thinking about retention, we're thinking now in terms of long-term retention. And you know, our feeling has been you know, perhaps retention is not best characterized over a 12-month period, you know, and maybe more of like an accent measure.

And then in terms of Mr. Thompson's question about how we define the medical visits. So, when we – the data that we used for our reliability and validity testing, it was data from the HIV Research Network and they collect data from you know, up to 18 sites. And this data is, you know, the original source is in the EHR or a paper chart and then they (abstract) it from either of those sources and put it into or upload to a database and how they have defined the medical visit as a face-to-face visit with a physician, a nurse practitioner, or a physician's assistant, you know, folks who are eligible you know, licensed to prescribe in their jurisdiction.

Now, in the event that hopefully we get endorsed, what we will do is we will specify these measures for use in electronic health records and we'll go back to the CPT codes and define which CPT codes would characterize a medical visit.

I hope this is helpful.

(Adam Thompson): Yes, that answered my question. Because I mean, how do we use these measures before I – I was almost certain that it would be a prescriber but because it didn't specify exactly, that was my question.

And also from a – just from a patient's viewpoint, I can say looking at the one-year versus the two-year difference, I know that it's a one-year mark – there is the scare of the initial diagnosis that sometimes keeps us in care. But it's the two to three-year time period that I see not only myself but the people in my community that tend to drop off because they get comfortable with their care.

So, from our viewpoint, from the community, I would say there is a meaningful difference in whether or not we make it through that two-year mark.

Marlene Matosky: And I think that's where we start talking about that access is perhaps that 12-month period, you know, that initial (inaudible) access and then going beyond that to the 24-month period. In addition with the gap measures, you know, we're looking at folks who are immediately, you know, in need of you know, visit, action, what-have-you and they may not be as Mr. Thompson is suggesting because they might be newly enrolled at the site or newly diagnosed, they might not actually fit it to that 24-month measure quite yet.

Reva Winkler: OK. All right. Great discussion.

Peter Havens: So if we agree – this is Peter Havens again that with the HRSA statement that retention in care over 24 months is perhaps most important, does that mean then finally that if – that we would potentially reconsider either the current one which has a 12-month timeframe or the new one which is suggested to occur over a 12-month timeframe.

Reva Winkler: Right. You will have the opportunity to do exactly that at the in-person meeting, to consider how these very much related measures you know, are some better than others and what's the best you know, group of measures? Do we need all four or they all four equally useful or is there a subset that would be more useful? So, yes, that is absolutely part of the steering committee's discussion and that will happen at the in-person meeting.

Peter Havens: And but for the HRSA person who just spoke, I guess I was interested to hear it from their perspective there was an important differentiation here about the date of the information that they would get or the you know, what they would know if they got one or the other of these included.

Marlene Matosky: Hi. This is Marlene again from HRSA. So, I'm not going to give you probably the most straightforward answer you want. However, the respond is that you know, we feel as though these measures are looking at perhaps slightly different concepts with the greater discrimination of the linkage –

access, linkage, and retention. I think these measures hit at slightly different things.

And you know, as we get to the next measure, we're looking specifically as just those folks who are newly enrolled which we know that there would tend to be "more vulnerable" to you know, lost the followup or lost the care within that first year. So, we see them all working together and constant in telling us slightly different things.

And I think going back to the previous conversation about the previous measure about the quality improvement, this gap measure is really one of those measures that's off the shelf, easy to utilize for quality improvement because the numerator population are those folks who actually you know, are needing followup, those folks who did not have that visit, that had that gap in care.

So, you know, I think that, you know, going back to what Reva had said, you know, I think you'll have definitely opportunity to you know, discuss this more (inaudible) what-have-you. And you know, as we get into that conversation perhaps on the 20th and 29th, you know, we have some thoughts about that also but we'll reserve our thought until that time.

Reva Winkler: Any other thoughts from the workgroup or should we go on to the scientific acceptability of this measure 2080? Dr. Farber, any thoughts?

Michael Farber: The group (inaudible) was high and I agreed with it too that there was a lot of – four out of four said that the scientific evidence is very strong and I think that that – it's somewhat of a – in this area again, a no-brainer because visits are important and also they're easy to measure because of coding issues. They're easy to find.

Reva Winkler: Any thoughts from anybody else on the reliability and validity? I just wanted to –just remind you that if the measures only had been tested at the level of measure score or the data element, you know, it can only be a moderate. But if there aren't any more thoughts there, how about usability and feasibility, I think these were some of the issues you were starting to touch on as you started comparing measures.

Michael Farber: Well, I think that the usability and the feasibility both had similar views from the group and that is three high and only one medium for both. I think that both issues on a measurement scale, visits are very easy thing to track and they're also very accurate to track because of the coding.

The only problem and what was mentioned is that visits are not an outcome. So, that we don't know what the quality of the visit is. So, we just have a visit and you know, they could vary a great deal but at least a collection of the visit, I thought there was a lot of consistency to that. And that's why usability and feasibility were both quite high on that.

Reva Winkler: Thoughts or comments from anybody else on the workgroup?

Adam Thompson: Yes, this is Adam. I have a question under usability and it was only based on the sites that they were citing as reporting on this measure from the Quality Improvement Project. They showed four data points, and the fourth one showed a sharp decline in the number of sites that were reporting. I know that also happens to matchup when some other big reports were due, and this would be a question to the measure stewards.

Was there any explanation as to why there was such a decline in the number of sites performing on this measure when the other measures were showing a consistent site reporting?

Marlene Matosky: This is Marlene Matosky from HRSA. We have found that it takes about a good three to four weeks to collect the data from other sites. And when we reported that data, I think we are about, you know, between one – week one and week two of the reporting cycle. So, it was just incomplete data for that time period and I apologize for not having noted that.

Adam Thompson: OK. No problem. Thanks.

Reva Winkler: OK. Any other thoughts on this measure or shall we move on to the next one? OK. So, the next measure is again a related measure, new measure from HRSA 2081, newly enrolled in medical care, the percentage of patients

regardless of age with a diagnosis of HIV who are newly enrolled and had a medical visit in each of the four-month period in the measurement year.

Dr. Farber, I think this yours again.

Michael Farber: Well, I think that this measure overall did not have the same degree of evidence that the other ones did. And I think it's reflected by what people commented on and also I think there was not as much studies and data there to compare this with.

So, if we went through it that the impact – well, let's see. The first one is …

Reva Winkler: Impact, yes.

Michael Farber: I think that we had only two that's out of the four that thought it had a strong impact. And then for performance, only one was high and one was medium. And I guess, two others weren't measured at that time.

Reva Winkler: Thoughts or comments from other workgroup members? Everybody (inaudible) the information that's summarized here pretty much reflects your general feeling as a group?

Male: One, I guess question is on the frequency of visits that form the numerator. And I'd be interested – this is a HRSA developed measure and the question is this is the – in every four-month's visit and one of the points that some papers make is that early on, more frequent visits might be needed and later, visits as infrequently as every six months are OK.

I'd be interested to hear from the developers what specific evidence they used to identify the four-month frequency of visits as an important measure of – of this adequacy of newly enrolled care.

Marlene Matosky: Again, this is Marlene Matosky from HRSA. I believe it was the paper that (Mike Mogavero) did back in 2009 where he looked at missed visits among establishing initial outpatient ambulatory medical care but quite frankly, I'm going to have to followup on that and double check because I don't want to

give the wrong information. But I think that was the article that suggested that that they're looking at every four months.

Adam Thompson: And this is Adam. One of the questions I had was around you're including in the denominator individuals who also transferred care. And if location has been stable, you know, receiving care for let's say, 10 years and they switch care, is the expectation of this measure with that individual would then have four medical hits over the next year if they had transferred it and what is the justification for that?

Marlene Matosky: So, the (inaudible) expectation is that with that person who has transferred would be included and it would be a visit every four months or it'd be three in a year. And what was the second part of your question, I'm sorry?

Adam Thompson: Just like – is there – is it shown that persons who transfer care who've been consistently in care with another provider who may just move geographically, do they need that space/time to show that linkage? Or my concern would be for instance, if someone was regularly in care and they switched doctors, and they're used to going once every six months because they're virally suppressed and adherent, that this measure would either low perform for the providers because the individual doesn't want to be seen that frequently or the patient might feel that there's a burden on their medical visits by having to go more frequently than they had thereby causing what would be like a part of a negative relationship that's formed, what their expectations as they come to see their doctor more?

Marlene Matosky: That's in interesting take on it. I don't think we necessarily came from that perspective. We came from the perspective of this is a person who's going to be mentoring potentially a new healthcare system, not just a new provider. So, you know, there was the potential risk that the client, you know, may have difficulty navigating a new system.

So, we didn't come from the perspective that you know, the client would be turned off by having, you know, had been – had a history of being you know, consistently adherent to visits and now having to, you know, (inaudible) to be a greater level of adherence to visits.

So, I think that's something that's interesting that I'm going to have to look into between now and the next time we talk.

Adam Thompson: Cool, thanks.

Reva Winkler: OK.

Michael Farber: This is Michael Farber. One thing I wanted to comment on based on what the previous speaker said, Adam, is that I think that newly enrolled, just as a measurement does have the ability to have a lot of variability because of you know, whenever you're starting new measurements, the early part of it sometimes can be difficult and that's what was mentioned.

In other words, of people transferring from other practices, so I think that inherently, this is a measure, although I think it's very important, it could have a problem with – with getting accurate information because it relates to people that are new into the program.

Adam Thompson: Yes, and I was specifically thinking about individuals who may have high co-pays with their insurance where this, you know, performance measure – if the providers really push it, could be an undue burden not on the system or the data but on the patients themselves.

Reva Winkler: OK. Anything else from the workgroup or do you want to talk about reliability and validity on this measure?

Michael Farber: Well, yes, again, this is a measure that the group as a whole felt it was less reliable and less valid by the numbers than the other two measures regarding retention. I don't have an explanation of how the others felt but I didn't feel there was enough studies to show this and I think that part of the problem also is the reliability of what you're actually measuring that it would reflect what you're trying to determine and that is that people are being seen early in the course of their identification.

Reva Winkler: Thoughts from other workgroup members? In general though, looks like everyone in the workgroup who submitted basically did feel that there was sufficient reliability and validity for the measure.

Michael Farber: Yes, I agree. It just wasn't as high in general as the other two.

Reva Winkler: OK. Do you want to talk about usability and feasibility?

Michael Farber: Yes, I think that those are also similar and that it is, I think the usability of it is – has been – is substantial because again, what you're measuring is something that can be found easily, and that's a visit.

And the feasibility, I think would be very high because it can be done and it has been performed. So, I think those two are high. The only, you know, really problem with this measure is how valid the numbers would be based on getting newly enrolled people.

Reva Winkler: Thoughts from any other workgroup members?

Adam Thompson: This is Adam. I would just ask the same question to Marlene, is the drop in performance sites again an incomplete data set?

Marlene Matosky: Yes, the same response as before. We were between weeks one and two in the data collection.

Adam Thompson: And then another question around feasibility. I know that we've spoken a little bit about incarceration before but I think specifically when looking at newly enrolled to care, is there any – was the exclusion the same thing the providers were just having difficulty with it because I would imagine, when looking at recidivism rates for incarcerated persons particularly persons of color in certain regions, that even individuals newly enrolled in care might dropout of care not because they don't want to be there but because they've been re-incarcerated.

And the inability to followup on that could be fairly high particularly for providers who might see patient populations with high rates of incarceration.

Marlene Matosky: So, this is Marlene again. It's not as though we don't think that incarceration has an impact of loss of followup or lack of retention. It's that we're thinking of down the road as we – as we move towards electronic health records and pulling the state out of electronic health records.

There are no structured data field by where, you know, we can track incarceration. So, this is, you know, kind of – I'm hesitant to open up that discussion that we had on the earlier measures but these are folks who are going to not make it into the numerator, so it's going to be that discussion as to why you know, are we only at as somebody said before 70 percent?

So, we're still looking at that 30 percent of patients and saying, you know, what it is about those 30 percent of folks that (inaudible) feel like that quality improvement piece would come in. So, if you have, you know, a proportion – a significant proportion that, you know, as you're suggesting is related to incarceration, then you would perhaps develop quality improvement specifically towards this.

Adam Thompson: So, you really envisioned then that the patients who don't show up in the numerator as (inaudible) in discussion that's happened at the clinical level?

Marlene Matosky: Yes, that's the, you know, where we really saw that the usability of this measure for the quality improvement piece and moving towards that realm and not being able to, you know, (clear) those folks out of the data. You know, one because it's important to keep them in but two is also, you know, incredibly difficult because there is no structured data fields for, you know, incarceration.

Adam Thompson: Yes, and as we move forward with that, again if you get endorsed here and move one, and one recommendation I would make is looking at (building) the implementation strategy around specifying that to people that the expectation would be that they look at incarceration rates in the persons who don't show up in the numerator. So, they have a little bit of guidance on that.

Marlene Matosky: Thank you. Noted.

Reva Winkler: OK. Anything else before we move on to the next measure? All right, so I think we need to go back to measure 403 which is the last measure of medical visits. This is the measure from NCQA on – that is currently endorsed. And that is the percentage of patients regardless of age with a diagnosis of

HIV/AIDS with at least two medical visits during the measurement year with a minimum of 90 or 180 days between each visit.

And I would note that this measure has two numerators so, the results would – there would be two actual results in the use of this measure. So, Dr. Havens, I think this one was yours?

Peter Havens: And so, you might get two answers for this, right? One was the 90 days and one was the 180 days?

Reva Winkler: That's the way I would read it. Does somebody from NCQA want to comment?

(Jenna): Yes, hi. This is (Jenna) from NCQA. Yes, the idea here is that there are two rates, so you will report both. It's not actually an "or" as part of one numerator. It's that there are two numerators. There are two rates.

Peter Havens: So, the first issue would be (inaudible) to measuring report and I think that for all the reasons we've talked about visit frequency is important to measuring reports. And so, this would meet those criteria. I think the definition of medical visit here as outlined is difficult for two reasons.

Number one that not all of the providers listed might be HIV providers. So, that saying that somebody who went to see their OB-GYN and taking that as a marker of retention of HIV care might oveimate retention in care. And the point was made by some of the other reviewers that data on staying in care aren't necessarily focused on seeing a provider as much as getting CD4 virus code on an every three or four months' basis and then the 180 days, and maybe an every six-month visit would be reasonable.

So, with those caveats, I think, the – we could say that the importance to measure and report is – has high impact and is based on reasonable quantity and quality of evidence.

Reva Winkler: Thoughts from anyone else?

Ed Septimus: This is Ed. I agree with the comment about defining what a medical visit is and by which provider. I think that's a – that's a pertinent comment to this measure.

Reva Winkler: OK. Thoughts from the measure developer?

Female: Hi. So, the reason why we have that definition and we do expand it to providers like OB-GYN is that there are areas in the country where there might not be HIV specialists. And so, the primary care for HIV patients is being provided by someone other than an HIV specialist, and it could be an OB-GYN, or it could be a primary care provider.

Ed Septimus: I understand that but the question is; what's the reason for the visit to that provider?

Female: That would be a very hard thing to measure. These measures were developed for a clean state program so, it's very difficult to say and to actually specify a measure that says that this has to be for these particular reasons. We can provide some guidance and I believe that's what we've done in the definition of the medical visit. But it's really more guidance that it would be a required part of the measure.

Ed Septimus: Let me go back to the validity of the measure. Would laboratory studies be a better idea of the visit? That could be captured.

Michael Farber: Again, I agree very much with that comment and that is that a lot of times that's how we measure things not by necessarily a visit you know, and medicate by what is entailed with the visit. So, with the OB-GYN doctor, since we don't know what the nature of the visit is, it would be interesting to compare the frequency of the CD4 counts and viral load in patients that are not seeing an HIV provider that might tell us what the nature of the visit is.

Adam Thompson: This is Adam. That was one of the issues I had with this measure was a lot of the evidence was around the importance of having those lab values yet from a patient's standpoint, I know that I can have those labs drawn and get those results without ever having to see a doctor.

And so it seemed like whereas the other three measures had evidence supporting the frequency and value of a medical visit, the evidence here was really around having those lab values. And so, it's sort of seemed it didn't agree. And to just to throw that in there too as we talk about it is also around the summary of data on disparities by population groups. It was a little concerning to me that it was considered an undue burden on data collection to collect basic demographics which to me would seem like it could happen on the first visit and would be in the system for every patient, regardless or should be.

Female: I can certainly speak to that. These measures were developed to be considered for use and the (inaudible) program. It's really not up to us to dictate whether or not they are reporting by disparities – you know, using disparities as a stratification option and they don't right now.

So, it's – we don't really have access to disparities data because we're not implementing measure, CMS is.

We agreed disparities are important and the way the measure is currently specified wouldn't prohibit a practice or groups to look at disparities if they were using this measure. We just don't have that data because it's only – it's being used in the (inaudible) program and they don't report using disparity.

Adam Thompson: OK. It just seems to me like it was being a little backward (inaudible) like the argument was being made that it was an undue burden at the data collection level. But it sounds like the persons who are implementing it aren't doing it, so therefore you're adjusting or not adjusting the measure based on the dominant implementation plan. Is that correct?

Female: I think – I think you've captured it correctly, yes.

Adam Thompson: OK. Thanks.

Peter Havens: So, this is Peter Havens. If I could then capture what I think I've heard is that people are worried that the data that are used as background information supporting the evidence to use this measure are inadequate because they are based on CD4, and virus loads, and monitoring frequency rather than on

physician visit frequency. Is that what I hear is the biggest criticism of the evidence base?

Adam Thompson: This is Adam. That would be – yes, absolutely. That's what I'm saying.

Reva Winkler: Thoughts from anyone else on the workgroup or shall we move on in the interest of time? All right. Dr. Havens, what are your thoughts on the scientific …

Peter Havens: Ms. Reva, I had some trouble getting off the mute but I wanted to say that the lab tests are an important measurement issue but they are not the sole issue for the visit, and some of those – the issues for the visit are really not easily measurable and that is counseling on safe sex practices, counseling on – and motivation for continuing therapy.

So, there's – there are many things that occur in a visit. Some of them are not easily measurable. The one that is very easily measurable is how often you get a CD4 count but that is certainly not the sole issue for being in the visit.

Female: Ms. Reva, on that topic I just want to remind you that one of the measures that (inaudible) has to evaluate if measure 404 CD4 cell counts performed. So, there is a measure around the laboratory testing.

RW: So, Dr. Havens, on this measure of medical visits, what are your thoughts on the reliability and validity of the measure?

Peter Havens: It seems like it would be relatively as reliable as in terms of being able to collect it. It sounds like it's mostly collected in CMS data but I think that electronic health records are in general able to collect this kind of thing as well. The 2079 medical visit frequency reliability data seem to look pretty good which would to me suggest that reliability data collected the same way with for this would look equally good.

Female: All right. Thoughts from anybody else in the workgroup?

Adam Thompson: This is Adam and I just have a question that may have been the misunderstanding sort of how things were presented. So it seemed as if the

validity testing was being presented for reliability testing and was been tested on electronic health record, get the data sources that were specified weren't EHRs. Did I – did I read that properly? So like administrative claims and – I got to find it here, it seemed like it was tested on something other than what they were saying their data source was.

Female: Comments (inaudible) to Q.A.?

(Jenna): This is (Jenna). I can jump in here. That is true. As I said, the measures were developed for the PQRS program which uses administrative claim and CPT category 2 code. The measures haven't been implemented in PQRS when they were tested back in 2009. So given that, the tested in the EHR should see whether that type of information is available even though the measures are using claims data and (CPT Q) code. And one reason why we think that works is that category Q codes actually – because they are quality claims or quality administrative data I'll say, they don't necessarily – they're not like other claims data. They could actually – provider could be using a paper medical record or an electronic medical record or some other types of claim to then report their category Q codes.

So we do think that the EHR testing data does support the category Q code specifications for the measure. I hope that clarifies. I know – I know it's a hard thing to understand.

Adam Thompson: I should thank you for answering and the second question is around face validity was sited yet there was no discussion of whether it was systematically assessed, where there's systematic assessment of the face validity?

Female: The – what we did is we had an expert panel that reassess these measures in 2012 and we basically surveyed them and gave them a 5-point record scale to grade the measures according to face validity.

Adam Thompson: OK thanks.

Male: So I had misunderstood then the electronic claims data so there is no estimate of the ability to collect this through the HR.

Female: I don't …

Male: And so this automated, calculated performance. I'm looking at 282.3, the measure of validity where it compared automated calculation and performance compared to manual calculation of performance than that automated is claims data not EHR based data.

Female: Oh no you're right. It is EHR data.

Male: Well I – then I'm really confused. So you can get it from EHR not just from clients?

(Jenna): That's right but there are not eMeasures. There aren't specifications for all of our HIV measures that we wanted – that we could submit with these other – this other specification. Now, I won't point out that actually for this measure and for the PCP prophylaxis measure. We are under contract with the ONC to develop eMeasures for consideration of use and the meaningful use program and so we do actually have eMeasures specifications available for this measure and for the PCP prophylaxis and we have gotten permission from ONC to share those with the stirring committee so we'll try to provide this to the steering committee before the meeting later in August and the measures were – there was some limited testing done as well that we will be able to include that testing data. And that's only for two of these HIV measures. It's not the entire (inaudible) of eight.

Female: Thanks (Jenna) that was very useful.

(Jenna): Sure.

Male: Except. So we can't grade the – so what's the answer to the validity question? Is it – has it been shown to be valid in an EHR based model? Yes?

(Jenna): Sorry, we were just discussing amongst ourselves.

We believe that they have – that the center has been yes.

Male: Then if I'm asked to interpret what's written that wasn't clear from – well was clear to Adam and then after Adam talked, it would include me and so we might think about making that clearer in what's written?

(Jenna): That's a good point meeting definitely take note of that and see if we can make it clear.

Male: OK.

(Jenna): All right. Any other thoughts on this measure on perhaps usability and feasibility?

Male: Well, for my perspective it's highly usable. We're not supposed to compare it to the two year measure which I also like as useful but I thought that the allowance per the 90-day and 180-day for the arguments we've just heard about lab testing versus visits together, I think makes it usable and this data – the CPT codes would be available in an EHR based data collection system so that enhances its feasibility as well.

(Jenna): Thought from anybody else in the workgroup?

Male: (Is that) my only thought is that I think if the face to face meeting of all of these either overlapping or similar measures, I think we need to discuss either harmonizing or combining some of these measures.

(Jenna): Actually it's already accounted for in the agenda because we will have a discussion of related measures like this and we definitely want you to have that. So we will be doing that.

Male: I think that'll – that'll I think help us a lot. We also – if we all can get together and maybe have a chart that kind of matches all these up, you've done this in other NQF subcommittees so we can see the different measures and exactly what the numerators and denominators and what the overlap is.

(Jenna): That's what we're intending to provide for you.

Male: Excellent.

(Jenna): All right. So now that we finished the visit measures, we have three remaining measures. The next one is a 405 PCP prophylaxis and I believe Dr. Havens set your measure as well.

Peter Havens: Absolutely and this is one of my favorite measures because it is so medically important so when think about the importance to measure and report, most everybody agreed that it had high impact could be used to identify a performance gap, there's a large body of evidence suggesting that this is useful as a measure of a health outcome. In the process of care utilizing PCP prophylaxis perhaps closer to a real patient level outcome in terms of its potential impact and so I think it scores highly in those importance and evidence.

(Jenna): All right.

Peter Havens: The scientific acceptability of the measure properties I think are also supported and I would score it high there most people agreed with that. It's usability and feasibility is terrible. This is incredibly difficult to reliably capture. It either takes a lot of programming which not everybody has available or hand counting because the denominator is so difficult and I noticed you've made it more difficult this year. Did you change the timing? So for the denominator exclusion something about the last three months of the year right?

So, it's a very – you have to be careful how you calculate it which makes it feasibility more difficult. But people do it I think. When you look at the how to do list, I'm trying to look for the denominator details page. You can see that a very complicated denominator to develop and …

Ed Septimus: Peter, this is (Ed) and maybe the developer can help us Because I'm looking at page 13 about the testing results and measure of validity. I think this kind of ties in to what you're saying and it looks like from the electronic medical record to visual of inspection of the medical records unless I'm reading this incorrectly that is fairly high degree of correlation with automated calculation as well as the manual calculation of performance so. It may not as difficult as it looks on the front end.

Peter Havens: Right but to do the automated calculation requires heavy programming in the background but not everybody would have potentially available.

Male: Can the developers comment on that?

(Jenna): Yes, we definitely can. It is – you did point out that we have made some changes to the measure since it was originally tested. This was to try and meet the measure more precise and to – one of the changes we made was to have that – to only look at patients in the first nine months of the year because we are allowing basically a three-month ran out after the test to allow for a second test and that second test is to make sure that the CD4 count going below that threshold is not just a (inaudible) that it is actually a sustained lower CD4 count.

Male: Right. From very supportive event, that's a very good change.

(Jenna): Thank you. This wasn't – this is another measure again was we created an eMeasure for it as part of this ONC – our ONC contract and we do have some testing data that will be able to share with the stirring committee on how feasible the measure was and to be honest to sites that we tested this measure in as an eMeasure, found that it was feasible because although it does have a number of complicated elements in it or has a number of elements which makes the complicated. There are generally elements that are found in structured field. So – although there's a lot of measure logic, it's actually a fairly feasible measure.

Male: Could do electronic without manual chart review?

(Jenna): All right. I do hear you point as a manual chart review. It becomes a much – much more complicated measure.

Male: Right. So that's my question. So when you say you've got an eMeasure does that mean that you have developed a program that anybody who's got (affect) can just get from you, download and it's good to go?

(Jenna): What it is is – when we have an eMeasure that we are creating for consideration of the use and meaningful use, we have – we do have the

measure in XML. That's what measure (implementer) could be able to review as the XML for the measure. You don't know if that quite answers your question?

Male: Well, so that – yes it does because what it means is then that it still going to require local programming capabilities and commitment to be able to make this into a feasible electronic measure? Because to get from an EHR to an XML data file requires some programming and these are important resource consideration for many programs.

Hello?

(Jenna): Yes. I hear that and note that …

Male: And so I mean I don't know. So it's – I think that with reliable as you point out if you got the data in an electronic health record and you program it together out, you can get it out. It looks great and it's a complicated algorithm but once you've done a programming, it works really nicely Because they're pretty strong data points. That's all great. The question of usability and feasibility, if we are going to say that this has to be adopted broadly then the cause of adopting it and the – how are we going to make that happen without an under burden to programs is an important consideration I think.

(Jenna): It is. It is certainly an important consideration. This is the difficult measure because the guidelines really do have. There are so many data points that might – that can get a patient into – into the denominator and for the numerator and then because the (inaudible) exclusion that does make it more complex. It is certainly a balance between making the measure as accurate as possible and aligning with the guidelines and creating something that's usable and feasible.

Male: So that's a crucially important measure because unlike visits to whoever it might be, this really allows you to assess the appropriateness of delivered care. So, it's a great measure.

Michael Farber: This is Michael Farber. I think that again as part of the discussion, the problem with the usability is the measure of CD4 count and you know that requires you the chart review or electronic records. But I think that you know if you have electronic records, the usability is very (inaudible). Without it, there is usability but as a high course because you have to do chart review. So in other words, if you were looking at just the initiation of treatment, well that the medication can be found on the claims and that are easy. But you wouldn't know that it was being used for the appropriate CD4 count so you do need – it does make in a more complicated measure but a very, very important one because of its success in providing long term survival quotations.

Male: And I have one other question, I know the HIV may have suggested getting rid of the less than 15 percent. So you've done that right?

(Jenna): I'm sorry. The …

Male: Well, you used to – the CDC recommended that you start pneumocystis prophylaxis for a CD4 cell count less than 200 or CD4 percentage less than 15 percent and in one of these things, I read some comments from the HIVMA that they like it but they wanted to get rid of the percentage less than 15 percent which you've done for people over age five but you've kept the percentage appropriately so for people between ages one and five. So …

(Jenna): Right. That's right. Yes. We do have – we are only using account for patient secures and all. They are using either the account or percentage for a patient one to five and can correct with the guidelines like you said.

Male: Which is appropriate right? Good.

Male: To make sure I …

Male: Oh go ahead.

Male: To make sure I understand that you are keeping it stratified by age then?

Male: Yes, you have to stratify by age. There's actually three age (grade at) one less 1 year, one 1 to 5 and one over age 5 and it was my understanding that the

HIVMA had wanted you to make the over age 5 only based on number not percent and – this looks like it's only based on number over age 5 which seems like you had responded to the HIVMA input and so I was saying how cool that was. I was trying to give you some positive feedback.

(Jenna): Great. Thank you.

OK. Any other comments on this measure before we move on?

Adam Thompson: Again, this is Adam. Just the same comment around disparity particularly when looking at the data that show that person of color is less likely to be prescribed PCP prophylaxis. I think it's important we capture that when looking at this because I agree with everyone that the prescription of PCP prophylaxis is crucial to long term survivor of those less living with the disease but particularly persons who are unlikely to receive it as a result of the arrangement this week.

Female: (But you) – this is what (Kelly and I) did? This is (Reggie's).

Female: Hi. Thanks Adam. All right.

Male: Very good point.

(Jenna): Let's move on to – we have two more measures. At 408 is a TB screening. This is a percentage of patients stage 3 months and older with the diagnosis of HIV/AIDS for whom there as a documentation that a TB screening test was performed and result interpreted for tuberculosis skin test at least once since the diagnosis of HIV/AIDS. And so (Ed), I think this one is yours, right?

Ed Septimus: Yes, and this was a very difficult one for me. I think other people had some similar reservations. Obviously, in terms of impact, TB and HIV is a bad combination. I went in a border state and I can tell that TB is alive and well in the state of Texas. So, I think the impact is certainly – certainly extremely high. The other thing that's important in this measure out is not a reliable predictor of increase risk of TB and that's correct and they present in different ways depending upon the CD4 count. (Partly, a) great opportunity for improvement.

According to the developer, we're talking about the mean rate of being only about 68 percent. (Inaudible) I don't do system standards for – as it's pointed out but what entity that is the physician, the group plan or the importer had to capture – had to capture this data.

So, I think it's a great opportunity for improvement. You can have a potentially a very, very high impact but if from a force with this particular measure is of course that whether it's a tuberculin skin test or whether it's using one of the new interferon gamma releasing assays (inaudible) of this in terms of predicting who's actually been infected and who should be on prophylaxis. Of course because of immunosuppression that goes down of course the guidelines say to be either/or. The interferon assays maybe slightly better in this area but – so but that's the easy part. I mean almost everybody agree that there should be screening and I don't think that's not where the difficulty comes from me.

So I think it's certainly the quantity is very high and I think looking at screen here in terms of how other people view that they read the evidence, in general would be pretty good and the healthcare outcomes to be good, the scientific acceptability was two to one. Where I think things begin to in my opinion, the potentially has to do …

(Jenna): Somebody has a lot of background noise. If you could put yourself on mute that make it easier for everyone to hear.

Male: Where things get to be difficult is in the reliability and validity and based on what was provided to us that when you look at electronic health records versus manual calculation, you can see there's a significant difference between the two. Very highly the labor intensive to capture this data and but having said that, there's still tremendous opportunities so that's why I think you see some changes in terms of the …

Female: (Inaudible).

Male: I think I'll sort of stop here and maybe ask the developers about their experience trying to capture this challenges here.

(Jenna): Great. This is (Jenna). Thanks for the opportunity.

Yes, so there is a bit of difference here between manual and automated calculation again because the measure is – the measures that we're submitting (inaudible) consider today is based on category Q code. The issue that really came up during testing was about having information for the measure and structure data feels – it's not that the information isn't available and a paper medical record for example. It's just that it's not making its way into structure data field for an EHR or in the EHR for an eMeasure.

So, we do think that even though there is a discrepancy between manual and automated calculation, the measure should still be feasible, valid, and reliable for using category Q codes because you can actually go to the paper medical record and find the information that you need.

Male: That is – anybody else – I mean Michael, Adam, anybody else want to comment on new experience with this?

Adam Thompson: Yes. This is Adam. My question was around the way the numerator is specified and the importance to measure in the evidence, you site the (inaudible) measure as evidence of its importance if the (inaudible) call measures specified differently and requires that the documents the way that's written is that maybe recorded like the result documented versus the result interpreted and having been in care myself and help a lot of other people through it. I think there's a significant difference between interpreted results and documented results particularly when patients are allowed to interpret the results themselves.

I feel like when you require results to be documented, there's more of ownership of that result on the provider side and it might solve that problem of providers both doing the TB skin testing, sending patients home. So my question is why is just results interpreted versus having those results documented?

Female: It was absolutely our intense for the results interpreted to be by a medical professional not by the patient. That's a good point you make that that is not

coming across in the measure specification so we can certainly clarify that because that was the expectation when we develop the measure. That was the intent.

Male: And I agree with that and obviously it's one of the advantage of using one of the releasing assays. It's not required to come back. Obviously, the results get in the medical record and the charge for that test is also probably available.

Male: (Inaudible).

Male: Go ahead.

Male: No, it's more expensive.

Male: Yes, that's I was about to say. You know its cost prohibitive for a lot of people but I've just seen too many people allowed to walk away and interpret them the result and the data on the validity of that result is really awful. So I think – you know if it was your intention that those results be documented and ready by medical professional, I would highly recommend that be written out.

Male: And I certainly concur with that.

Male: I entirely agree with that comment and that is that if we want to see accuracy or at least better accuracy with the tuberculin skin test that it should be done by someone who is experienced and I think that's the general problem I even felt with this in general and that is that there is a lot of variability even among people who are experienced. I don't think that we can eliminate that by this measure but I think requiring a health professional to interpret it and not the patient would be at least a start.

Male: And remember the cut off for this patient population is 5 mm.

Steve Brotman: Hi. This is Steve Brotman. You know even though it's a health professional I need to clarify that it should be in a health professional that is trained. I mean I've seen a lot of nurses and other prior professionals interpret these things any type of training whatsoever and we had to (inaudible) for interferon testing.

Female: Hi Dr. Brotman. And how are you? Glad you can join us.

Steve Brotman: Thank you so much.

Female: Lurking in the background there.

Male: This gets to some of the discomfort I have with this measure plus the fact the predictability of test depending upon the level if immunosuppression of course is as everybody knows is not as predictive as let's say CD4 count for PCP prophylaxis.

Female: OK. Any other thoughts for today? Any other questions for the developer or clarifications you'd want that will help your discussion at the in-person meeting?

Female: (Inaudible) there's one more right?

Female: Right and we have one more measure.

Male: (Inaudible) I think this is – this is a measure at least from my perspective, of all the measures we've looked at that I have the most discomfort about whether or not it's suitable because of all the reason could've been mentioned.

Adam Thompson: This is Adam, the only other thing I would add is looking at the fact that you know TB is one of those things that's environmentally acquired not behaviorally and so the inability to look at this measure as far as persons who are incarcerated, persons who experience one of the – sort of spectrum of homelessness or in certain institutional setting. It really limits the ability to target down where the TB is coming from I think at the quality level.

Male: Great point.

Female: OK. We have one final measure to do as (inaudible) have mentioned and our time is running short so let's hear your thoughts on measure 409. This is actually transmitted diseases, screening for chlamydia, gonorrhea, and syphilis. This measure actually was endorsed as two measures and with the – that have not been combined into the three STDs and so it's the percentage of patients age 13 years and older with the diagnosis of HIV/AIDS who have

received chlamydia, gonorrhea, and syphilis screening at least once since the diagnosis of HIV infection.

So, (inaudible) I believe this one's yours.

Female: Yes it is and I'm going to talk based on the document that you sent me last night. I don't know if more people have responded to your survey at this point because I see when they (inaudible) but …

So based on what I see, there was consistency and (inaudible) but the biggest issue that was brought up was one with disparity and lack of classification of the data although the (inaudible) did mention that there is a disparity. There was less justification of why they cannot be certified. And the second major issue that came up was about the glitch that's mentioned in the validity section where the EHR data inconsistency was closer to 30 persons if I'm right. Its 33 persons and those are the two main issues that came up. (Inaudible), do you want me to go by section by section or …

Female: No why don't you hit the high points for each section.

Female: Sure. So the impact as I mentioned, the main question impact for one of the (inaudible) was about this classification. On that it was consistently seen as high impact. The evidence (inaudible) with what the measure (inaudible) noted except for one question, one of the respondents having issue noted as incomplete or incomplete for review and measurement and actually it would be nice to hear from the person who noted that.

Adam Thompson: Yes, that would be – this is Adam, and that just had to do with the fact that the discussion didn't seem as robust as some of the other measures and the way it was written as well as the NQF endorsement material guidance on us. It seemed like it required more of a discussion around what exactly the studies around the quality, quantity, and consistency were. So, it wasn't that I necessarily disagreed. It was just based on how it was presented on the paper that it didn't necessarily need all those requirement.

Female: Any comments from the measure that was (inaudible) or shall we move on (inaudible)?

(Jenna): This is (Jenna). Thank you for noting that. We can certainly go back and look and that and see if there's a way for us to strengthen that section.

Female: Thank you. Moving on to – (slide number) one to the next section.

Female: Sure.

Female: Moving on to the scientific acceptability data liability and validity was consistent with what the – as a group, it was consistent except for again incomplete to the few. One of the point that have brought up is about the glitch that mentioned and the reliability and the validity section for 30 percent difference between the automated and manual. And that makes it difficult to (judge) the reliability section. I actually would like to get a discussion from (inaudible) on that?

Female: And maybe it would help for me to clarify (inaudible) and certainly if others want to have a discussion about the …

Male: Well, I think the other thing is – I think was implied is, it says there that sometimes, these were done but for some reason, the electronic record did not capture it and you talk about the technical glitch.

Female: That's correct. Yes. The – what happens here is that at the particular site where the testing was being done, they actually had a problem in their EHR where test data was not going to the right standardize deals and the EHR so the information was there. It just was in the wrong place which means the automated calculation not correct. But the information is actually available in the record.

Male: (Abby), have you looked other electronic medical records to see if they could be more accurately captured?

(Abby): We have not tested it in other medical records. No.

Male: I think that may be important going forward especially with the (inaudible) care acting incentives to put in an electronic medical record. We're going to see more and more of this and the ease of trying to capture some of these

measures if it can be automated who will put a tremendous – who will make it the burden much less on the provider side.

Female: OK. Moving on to the usability it was consistent together that in (inaudible) usable and the measure was using the existing mechanism to get the data with the CMS PQRS was currently – measures currently used in that and that was mentioned in the usability section and again some of the (inaudible) glitch here than as mentioned that should be (six) before we could see more data on it and the feasibility was similar to the other proof that the data certainly can be collected during and quoted during the – someone gets care in the medical setting.

There was concern over screening with (inaudible) test in difference between the chart and electronic data. I am actually not the right person to talk about so I would like to turn to the person who raised this issue.

Adam Thompson: Yes, this is Adam, the reason I bring that up it has to do with by combining these two measures, my concern is I've seen now in the field that some providers interprets screening to be screening for sexual activity and if sexual activity is not indicated, then the test is not performed and when we try to do this in Virginia, we need a statewide project.

We interviewed providers all over and everyone was interpreting screening to be different and so we actually had to specify to them that it was a serologic test expected for syphilis and because we're talking about HIV, I think this is how people end up not being screen for HIV as well.

And so my concern is the way the numerator is specified that it shouldn't say necessarily whether a patient was screened but that they actually received a test just so that when it's been implemented, providers are clear that it's not screening for sexual history indicating a test but is actually the provision of the serologic test itself.

Female: We absolutely – we agreed that what we mean here is that they are getting a laboratory test for gonorrhea, chlamydia, or syphilis and if that's unclear in the measure specification, we're happy to clarify it but we do mean a lot and not screening for behavior.

Adam Thompson: Yes. I would just hate to see that the higher rates of gonorrhea and chlamydia testing go down because people are lining it up more with screening the way they do syphilis and people seemed to be more concerned about the outcome of syphilis in an HIV patient. That's my – my only concern.

Male: If the developer accept it more specific language for screening?

(Jenna): This is (Jenna). Yes. Like I said, just a second ago, we're happy to clarify that we do mean an actual laboratory test not screening for behavior.

Male: Yes. And the same thing could be said for chlamydia and gonorrhea too.

(Jenna): Right. It would go in there for all the lab test.

Male: Excellent.

Adam Thompson: And this is Adam. I also had a concern around it only being done – performed once in HIV diagnosis. I mean if an individual is HIV positive and they did not acquire the virus either through vertical transmission or injection drug use, I think it's safer to assume that the sexual activity will continue.

I mean we don't become asexual the minute we're diagnosed and so I'm concerned that there's a lot undiagnosed gonorrhea, chlamydia, and syphilis. If there were only performed once because I would assume most individuals would do that upon intake and the care, check off the box, and they never think about it again. So it was just a concern I had as far as the frequency of this testing done particularly looking at the data on transmission, we know that these individuals were positive or continuing to engage in sexual activity or we would not have higher incidence.

Female: We did have quite a bit of discussion about this at our expert panel and we have one of our experts on the line today so Dr. Michael Horberg, if he's able to speak up, I certainly welcome him to – the issue is that the guidelines are fairly clear that you should only be doing annual training for this other STDs if the patient is sexually active and while many patients with HIV might be sexually active, there are some and are – and a few of our expert has support

this that are not. And that doing an annual screening is really overused for those – for that particular population.

So, it wasn't – it's not very clear cut but I think if you do look at the guidelines, they do – they do say that you should really only be doing the annual screening for patients that are sexually active. And as you all can imagine, identifying sexually active patients using claims data is difficult. We do (inaudible) actually does have another measure where we do that but women, we use things like pregnancy test, STD test and things like that to identify sexually active women. It's harder for men but it's very hard to do using claims and isn't particularly precise.

Male: Well let me ask you these I mean, let me ask a question a different way, what percentage of people that are in this category are not sexually active?

Female: I don't have data available at my fingertips right now to speak to that. Again, I don't know if anyone else here would know that. I don't know that there's published data about that but we could look.

Male: Yes. Because my concern is as was stated is that we're allowing a pretty big hole here for people just to check off the opt out that it didn't screen because they didn't think their patients were sexually active if the majority or vast majority of them still are.

Female: Understandable concern and like we can look to see if there is any data. I don't know if there's published date about that but we can look.

Male: And there are other people not on the A-panel but on the big panel that have a lot of experience in this area as well. So perhaps, when we meet face to face we can have a more full discussion about that. As I – and I agree I think this probably in most patients should be yearly.

Male: And also we might be able to just look and see what percentage of patients was screened for sexual activity because that might show why it is not happening as well.

Reva Winkler: OK. Any other thoughts on this measures as we're approaching the end of the call from anybody in the workgroup?

OK. Operator, do we have anybody from the public listening in who'd like to make a comment?

Operator: At this time, if you would like to make a question or a comment, please press star then the number one on your telephone keypad.

And you do have a question form Michael Horberg.

Reva Winkler: OK.

Michael Horberg: Hi. This is Michael Horberg, Chair elect of the HIV Medicine Association and third on the NCQA committee here. Sorry for some reason, I couldn't comment earlier. The big question on this last one, I think it's important to bring out on the question of feasibility. Looking at a lot of large electronic health record, electronic health system, and manual charting, it's very difficult to discern who is and who is not sexually active.

So, effectively, the only way to make this feasible and operationalize is to then set screening everyone annually but that's actually a waste of cost so that was why it was determined to do it as a one-time metric. Assuming you had better data systems and you could key better data system and I'm not even sure that the (Inaudible) clinics can annually check this either.

It was looked to be very wasteful and I want to make one other comment that I have your attention and really urge the committee to – when it comes to the medical visit metric to either synthesize or call them down because as standing now, all the metrics make it very difficult. But I would note that getting an office visit, a good medical visit is more than just CD4 count and viral load and then for some reason that was almost (inaudible) in one of the comments earlier today.

Reva Winkler: Any other comments operator?

Operator: You have no further (inaudible) questions.

Reva Winkler: All right. OK.

Well I want to thank everybody on the workgroup and if you know Steve Brotman who is the other Co-Chair was listening in on the background, I don't know if all of these steering committee members are invited to join in in this workgroup calls so I know your time is limited so I appreciate the time you have invested with us today as well as doing the evaluations.

Basically, the in-person meeting we will be going through these measures in a similar fashion but it will be a little more structured because – for each of the criteria, we will ask the committee to vote on whether you feel it meets the criteria and we have these little electronic vote apparatus that we'll be using to collect your vote.

So, this sort is a preview of what to expect from the in-person meeting but we'll be doing it on 27 measures.

So, (Alexis) is there anything else you wanted to tell everybody in preparation for our in-person meeting which is going to occur less than two weeks?

(Alexis): Right, so next week, we will be sending you the final agenda for the in-person meeting as well as the summaries from these workgroup conference calls. As the committee will receive all four workgroup preliminary evals as well as the summary from each of the four calls.

Ed Septimus: And this is Ed just to tell the folks who are new. I think we're probably going to need to carve out some time in the second day to see if we can combine and harmonized some of these measures. So, I'm pretty sure that's going to be on the final agenda.

Reva Winkler: Right. It already is Ed so we anticipate that. That's a very common part of NQF projects so we certainly will be wanting to do that and we'll provide as many comparative tools as we can to assist in your work.

Does any – in the last couple minutes, does anybody have any questions in terms of expectations for the upcoming meeting or the evaluation of the measure or anything else we can help you with?

OK, well it doesn't sound like it. So again I want to thank you all very much. I think the evaluations were very thoughtful. Your discussion was wonderful. I think you will – are prepared to share your thoughts with the entire committee and hear even other input as the entire committee goes through their evaluation.

So, we really look forward to meeting all of you in person in two weeks in Washington and at any time, feel free to contact myself, (Alexis), or (Adele) if you have any questions whatsoever, we're here to help you and we do thank you very, very much for the work you've done on our behalf.

So that's all for today. Thank you and have a good evening.

Male: Thank you.

Male: Thank you.

Male: Thanks.

Operator: Ladies and gentlemen, this does concludes today's conference call. You may now disconnect.

END