NATIONAL QUALITY FORUM

Moderator: Board of Director's
February 28, 2017
11:00 a.m. ET

Operator: This is Conference #: 92637110.

Christy Skipper: Good morning, everyone, and welcome to the first workgroup call of the Infectious Disease Standing Committee. My name is Christy Skipper and I'm the Project Manager. And I'll turn it over to my colleagues just to introduce themselves.

Mauricio Menendez: Good morning. Mauricio Menendez, I'm the Project Analyst.

Melissa Marinelarena: Good morning, everyone. This is Melissa Marinelarena, the Senior Director on the project and thank you for joining us this morning.

Christy Skipper: Thank you. So the purpose of this call today is to allow the standing committee members to have preliminary discussions about the measures that you evaluate at the in-person meeting which is actually about two weeks away, two weeks away.

Many of you are new to the NQF process so this is an additional opportunity to ask questions about the criteria, expectations of U.S. committee members and then, also, we have the developers on the line to answer any questions or clarify anything that has given you pause as you sort of walk through your measures.

So before I go any further, I'll turn it over to Mauricio Menendez just to take a roll.

Mauricio Menendez: Please say here when your name is called. Woody Eisenberg?

Woody Eisenberg: Here.

Mauricio Menendez: Adam Thompson?

Adam Thompson: Here.

Mauricio Menendez: Emily Aaronson? Amesh Adalja? Esther Babady? Nanette Benbow? Kathleen Brady?

Kathleen Brady: Here.

Mauricio Menendez: Laura Evans?

Laura Evans: Here.

Mauricio Menendez: Piero Garzaro?

Piero Garzaro: Here.

Mauricio Menendez: Donald Goldmann? Jeffrey Hart?

Jeffrey Hart: Here.

Mauricio Menendez: Michael Lane? Jeffrey Lewis? Melinda Neuhauser?

Nanette Benbow: Nanette Benbow.

Mauricio Menendez: OK. I got you. Rocco Orlando? Jamie Roney? And Pranavi Sreeramoju?

Christy Skipper: OK.

Mauricio Menendez: OK, thanks.

Christy Skipper: Thank you. And then, I like to here, is there any representatives from HRSA on the call and if you could just acknowledge your presence or say your name.

(Marlin Matoski): Yes. Good morning, everyone. This is (Marlin Matoski). I'm also joined by one of our health scientist, Meredith Brantley and also our ECQM experts, (Laura Monroe), (Renee Rickwood) and (Ruth Martin).

Christy Skipper: Thank you. And I just want to remind everyone to – that you need to be logged in to the webinar and dialed in on the phone in order to fully participate in the discussion today.

So before we get started, I just want to give you an overview of how the discussion will flow. So we'll all be following along on the measure worksheets, so in a few minutes we'll get started with measure 2079 and you'll see that on your screen once we get to it.

But we'll walk through the measure worksheets and we'll call on the lead discussant to give a real quick summary or definition of the measure and what it is it's suppose to do. And then, we'll walk through each measure criteria by criteria giving the lead discussant a chance to give any initial questions or comments or ask questions of the developer, and then we'll turn it over to the committee to provide any questions or comments. And then, we'll move on to the next criteria and that will be the way that will flow for each measure.

So we'll go ahead and get started. Well, before we get started, are there any questions about that process.

Donald Goldmann: Yes. So this is Don Goldman and I have a question in general. I'm in workgroup 2 but the invite I have was the group 1 call at 11:00. My standing is the group 2 call is 1, is that right?

Christy Skipper: The group 2 call is actually tomorrow from 1:00 to 3:00 p.m. Eastern but we did invite all committee members to attend both workgroup calls just to give measures. But we would definitely, if you're available to attend workgroup 2 calls since you are assigned to that one to call in and participate in that call tomorrow.

Donald Goldmann: Oh yes. So that's the one I can't attend because it's during our board of trustee's meeting. That's the issue.

Christy Skipper: OK. Well, we would welcome you to stay and here about the HIV measures today and provide any input or thoughts that you have.

Donald Goldmann: All right. Maybe I'd learn the process this way but I won't be able to be on the one tomorrow obviously, thank you.

Christy Skipper: All right, thank you. Any other questions or comment before we begin? All right.

Well, we'll start off with measure 2079, HIV Medical Visit Frequency. And our lead discussants are Nanette Benbow and Kathleen Brady. So again, if you could just start off talking about what the measure is, the rationale. And then, we'll go on through criteria by criteria. And you can decide them on yourselves who wants to start off the discussion.

Kathleen Brady: Nanette will be …

Christy Skipper: Kathleen.

Kathleen Brady: Yes. Yes. You want to be start?

Christy Skipper: I'm volunteering you, yes.

Kathleen Brady: OK. I've done this before but it's been awhile. And this is not a measure I've reviewed individually before. But anyway, so the title of the measure is HIV Medical Visit Frequency, was developed by HRSA. And basically, the brief description is the percentage of patients regardless of age who have a diagnosis of HIV who have at least one medical visit in each month period of the 24-month measurement period, with the minimum of 60 days between each medical visit. And so, medical visit has to be in an outpatient ambulatory care setting with basically someone who has prescribing privileges, whether it's a nurse practitioner, physician and/or a P.A., who provide comprehensive HIV care.

And the rationale for this one really is based upon, it's a process measure but it's upon that for retention and care, especially in the first year outpatient care is associated with outcomes in terms of receipt of antiretroviral therapy and

delayed time to virologic suppression, and people newly diagnosed with HIV. And so, that's the majority of the data that is presented.

And there is – so basically there's also information about how no show clinic visits have conveyed a 17 percent increase risk of delayed viral load suppression in some of those studies. And some information about the – how retention in care is a vital measure in the HIV care continuum which was started to be – used really in about 2011, and really outlines the steps that are required in medical care of people with HIV from time of diagnosis, linkage to care, retention of care, receipt of any retroviral therapy and substantially viral suppression.

And that model has been incorporated into the National HIV/AIDS Strategy where there are outcome measures listed for all of those indicators and basically they provide the most recent nationwide data from CDC from 2014 that indicate that 86 percent of people living with HIV have been diagnosed. Only 40 percent are engage in care, but that actually, that data is not based on this measure, 37 percent have been prescribed ART and 30 percent have achieved viral suppression. And so, this is an opportunity for – to this measure to try any increase retention and care which hopefully will lead to improvement and viral suppression. I think that's it, it terms of the basic rationale.

Christy Skipper: Thank you. That was Nanette, right?

Kathleen Brady: No, that was Kathleen.

Christy Skipper: Yes. Kathleen, sorry. Nanette, is there anything that you would like to add?

Nanette Benbow: No, I think she covered it.

Christy Skipper: OK. Do you want to move on to talking about the evidence?

Nanette Benbow: Yes, that would be fine.

Christy Skipper: And if you have any questions that does need to most of you, and if you have any questions regarding what the staff roll, that's what we are here for as well.

Nanette Benbow: OK.

Kathleen Brady: Yes. So in terms of the evidence, I would agree based on the algorithm that the information provided is the – for the evidence is insufficient and that is largely because there was not a systematic review of the evidence specific to these measure. And however, some of the evidence that was submitted is graded but based on the fact that it's not a systematic review of evidence, it gets a rating of insufficient.

Do you want me to go over the evidence that was submitted?

(Off-Mic)

Christy Skipper: Go ahead.

Christy Skipper: It is a maintenance measure and so, there is less impetus on the evidence unless there is new information or change in evidence since the prior evaluation. So that's the difference with the – between the maintenance and the new measure is that the main focus really is on performance gap. If you also like to know new evidence about this measure, and it doesn't need to discuss then we can move on to performance gap.

Kathleen Brady: Do you want me to …

Nanette Benbow: I'm fine with that I didn't – I'm not aware of any new evidence for the measure.

Kathleen Brady: No, neither am I.

Melissa Marinelarena: And we believe, (Marlin), if you can confirm this. The evidence that you provided to us, is just updated, the guidelines have been updated, correct?

(Marlin Matoski): Yes.

Christy Skipper: Thank you. OK. So now, if there are – if any committee members on the phone have any questions or comments about evidence or – then you can share those now or we could move on to the performance gap.

Woody Eisenberg: This is Woody Eisenberg. I do have a question. Is there any evidence linking this every six-month periodicity, retaining of member of patients to any of the outcomes that will be reviewing either in this measures or in any of the ones, the rest of the morning?

Kathleen Brady: The specific measure.

Woody Eisenberg: Pardon me?

Kathleen Brady: I ask if you meant the specific measure?

Woody Eisenberg: Yes. But, so then the measures have so many common elements to them that I think your answer will be helpful going forward to. But for now the question is, I know the six months is convenient and it make sense because we live 12-month years, but I'm wondering if there's any evidence to actually support a visit every six months.

Kathleen Brady: That's based upon what in the guideline where, I mean, that basically people who have, you know, who are suppressed. So can – so basically, it's specific to guidelines is the patient on a stable suppressive (ARV) regimens, viral load should be repeated every three to four months that had a grade of (A3), or as clinically indicated to confirm continuous viral suppression. And clinicians may extend the intervals to six-month for adherent patients who's viral load has been suppressed for more than two years and who's clinical and immunologic status is stable and that's also an (A3) recommendation.

Woody Eisenberg: Very good.

Kathleen Brady: So it's based on those guidelines.

Woody Eisenberg: Thank you.

Adam Thompson: And this is Adam. I might have a follow-up question to that. Well, I see the guideline there but I think it's possible also the way folks run now with laboratory testing, what they're asking for here is a visit with a prescriber. I

can have my viral load gone through standing order and not actually see a prescriber.

And it seems to me, a follow-up question that this – all the evidence we've seen is about how important this retention is in the first year, but I don't see anything about it in the successive years. And when you look at the guidelines, while they recommend two viral loads, we only recommend one medical visits for folks who are durable suppress, and four visits for folks that are not suppressed. So it seems like confusing to me around whether the measure meet sort of aligns with any of those guidelines.

Christy Skipper: Would the developer like to response to that?

(Marlin Matoski): Sure. I was waiting to figure out if Kathleen was going to response. So thank you, Adam, for that comment. As you had pointed out, if somebody is durably virally suppressed for multiple years, based on just their viral load alone, the physician or the practitioner may choose to do less visits. However, I want to make two points.

One is that, in the context of HIV medical care, we're not only dealing with folks who have HIV. We're also dealing with all of their other comorbidities. And unfortunately, we do recognize that people living with HIV have a significant number of other comorbidities. So they may be coming in to address other items as well, more frequently than just that once a year if they are in fact virally suppressed. That's number one.

And then, number two, although these measures are applicable to any provider in this country providing HIV care, inclusive and exclusive of those funded through the Ryan White Program, we did look at Ryan White did, as we think we have one at the largest data sets in the world around people living with HIV.

And in fact, we do see a statistically significant and clinically significant difference, and the rates of viral suppression between the clients who have only one visit in a 12-month period, compared to those who have two visits, three visits and beyond. Now, the effect of the impact of visits diminishes as

we get more and more visits, but we do know that, you know, we're looking out at least, the good 10 percent, 12 percent difference on the aggregate between viral suppression between those who had one visit and two visits. And then, when we stratify to look at very vulnerable populations. We see even greater differences in viral suppression.

Now mind you, we haven't published this data. This is very, very recent data based on our 2015 data that which I just collected and released. But we are seeing some differences. And in fact, Dr. Cheever, Dr. Laura Cheever who is our Associate Administrator here at the HIV/AIDS Bureau, who also sits on the HIV HHS Guidelines Committee will be sharing this information with the guidelines committee to hopefully provide some greater detail with respect to, you know, the timing of visits and such.

Kathleen Brady: Yes. And in Philadelphia, we also have some local data that also supports that long term retention in care is predictive of durable viral suppression as well, and that's looking over a five-year period. But we also actually, I have some data from a student on my desk. And, you know, retention in care, I mean, I think as (Marlin) mentioned is not just important for HIV but it is very important, you know, for other comorbidity. And in an analysis that we just did a 2015 data, retention in care is significantly associated with screening for hepatitis C, as well as being in care for one hepatitis C, and also with evidence of effective treatment for hepatitis C whereas viral suppression is actually only associated with screening and it is not associated with being in care for one hepatitis C or with resolution of hepatitis C after treatment.

Adam Thompson: So this is Adam. I just have a clarifying question to NQF. As the committee, do we consider evidence that's not presented before us are published, how do we incorporate that?

Melissa Marinelarena: You can if there's additional evidence, we can ask the developer to provide it to us. They can send it to us if it is available prior to the meeting, then we can get it to you before then. If the committee has evidence that wasn't provided by the measure developer, if you send it to us, we will disseminate it to the rest of the committee as well. I mean, you can take that into considerations.

Adam Thompson: So I think my follow-up question was then to the developer, if it's possible for us to have the data that you were speaking about to review at our in-person meetings. Because I think, if those data, you know, are doing what you say they're doing, I think it makes the argument stronger.

(Marlin Matoski): Sure. I will work on succinctfully writing that up and sharing it with you all.

Adam Thompson: Thanks.

Christy Skipper: OK. If there are no additional questions regarding evidence, we'll move on to gap. And so, Nanette or Kathleen?

Kathleen Brady: So the developer has presented data that is summarized in the measure worksheet. I'm trying to get back to my screen. There it is. So – and they present data from the HIV research network, which for folks who don't know is consortium between community and academic site that provide HIV care linked by centralized data coordinating center. And so the HIV (RM) is composed of 11 sites representing four major geographical divisions. And basically, they did not have data for 2011 to 2013 but they otherwise present data for 2007-2008, 2008-2009, 2009-2010 and then 2014-2015. And they showed the mean performance rate was 66.7 percent in 2007-2008 and increased to 72.6 percent in 2014-2015.

And then, they also presented client level performance scores for each of the medical visit frequency over the – for a time period and basically it does indeed show that there are some disparities in HIV medical visit frequency among Hispanic males, transgender and persons who are aged 18 to 29. In the additional information, there was also, I think, some data that was presented for the entire time period by facility that also showed that there was significant variability by each of the 11 clinics.

Christy Skipper: OK. Would the co-discussant like to add any thing to it or are there questions or comments from the committee?

Nanette Benbow: No, I think, I'm fine. I mean, I would just add, I think Kathleen said it, but it's the site in the HIV (RM) do seem to be sort of by their selection representative

of different geography insurance status and coverage types of people living with HIV.

Christy Skipper: OK, thank you. So now moving on to reliability. Any comments, questions regarding the reliability specifications, testing and the same for validity and validity testing?

(Off-Mic)

Kathleen Brady: Hold on. Yes. So, I guess, I mean, I found that all the data elements were clearly defined and the appropriate codes were included in the spreadsheets for the types of visits that could be included. The logic are in the calculation algorithms seems clear. We actually used this at a local level with our CAREWare data and don't have any problems calculating this measure. And so, I think it can be consistently implemented based on the data elements that were provided.

In terms of the testing that basically, they estimated reliability using a beta-binomial mode to assess the signal to noise ratio and they found a median reliability of 0.97 which is very high. And that between clinic variance was low at 0.0072. So based on that, it needs the reliability algorithm, the data presented is high. And then you ask about the validity, basically the validity is based only on face validity because it was – it was established using a technical advisory panel. And I think one of my only questions with regards of this was, since this measure was first approved in 2013, I was wondering why they had not updated the validity testing.

Christy Skipper: Is that a question for the developer?

Kathleen Brady: Yes.

(Marlin Matoski): This is (Marlin). So, as you probably saw from the subsequent measures and which we'll be reviewing. We have pivoted from focusing wholly on the chart abstracted or paper measures to now focusing much of our efforts in the ECQM development. And our understanding from NQF was that if the face validity or the validity was deemed acceptable with the initial review, we wouldn't have to invest more time and effort in working on validity.

Melissa Marinelarena: Hi, this is Melissa. That's right. So just to explain again, the ratings of insufficient and especially for face validity, it's only because we have a very specific definition for face validity which we included in all of the preliminary analyses.

So, what we do is we're not saying that it's not good. We'll just ask the committee that based on the information that the developer provided, do you believe that this, you know, that that information that the measure score does distinguish some good quality and poor quality. And that's the only reason why, is because face validity is the very minimum that we request and we do require very specific language. And then if it's not in there, we look to the committee who are the experts and you have used the measure to determine whether it's sufficient or not. That's all about this.

Kathleen Brady: OK. And I think the last part of the validity is about exclusion so patients are excluded from the measure if they die during the measurement period. But, actually I got it based on the other one I know the mortality information. It's represented a small number of people who end up excluded. It also I believe, excludes people who diagnosed with HIV in the first three months of the 24-month period, is that correct, (Marlin)?

(Marlin Matoski): Yes.

Kathleen Brady: But how come that is not listed here?

(Marlin Matoski): That's not actually an exclusion, that’s more – that's a denominator criteria.

Kathleen Brady: OK. OK. That's fine.

Woody Eisenberg: This is Wood Woody Eisenberg. I have a question for (Marlin). (Marlin), has there been any correlation between this measure and viral load in your patients?

(Marlin Matoski): So we have looked at in the national data set as well as other folks in the university. I've looked at – there is a correlation between greater retention and greater viral suppression. And in fact CDC has actually written a paper. I

think it came out in JAMA about two years or so ago. And they looked at the patients who are not virally suppressed and it turns out that of those who are not virally suppressed, a significant portion of them have not been retaining care and it's that combination of not retaining care, not virally suppressed and they're also contributing to, I think, two-thirds of the new diagnosis. So it's sort of one of these things of, you know, chicken or egg sort of scenario, in terms of the relationship between retention and suppression and then also ultimately new diagnosis.

Woody Eisenberg: And can you tell us how CDC is defining retention?

(Marlin Matoski): I believe the CDC was looking at two visits in a 12-month period.

Woody Eisenberg: Thank you.

Kathleen Brady: It's not necessarily a visit. It's typically, since, for CDC, it's prevalent technology. It's defined through laboratory data. So it is evidence of two or more CD4 counts and/or viral loads at least 90 days apart within a calendar year.

Woody Eisenberg: Right.

(Marlin Matoski): Thank you.

Woody Eisenberg: Thank you.

Adam Thompson: And this is Adam. Just a follow-up to that last question, can you speak to us all, we would talk a lot about, so far looking at the evidence that retention has been tied to suppression, are you seeing on the back end that there are significant number of individuals who are suppressed that don't need the retention measure to kind of get a sense of whether there's a log on both sides?

(Marlin Matoski): So, sorry, I was having, I was thinking through something there. So, we have not investigated to your point, Adam, those who are virally suppressed not meeting the 24-month measure. At this – excuse me, at this moment, we're not able to link the Ryan White HIV/AIDS Program data across multiple years.

However, there is a correlation between when the HIV research network has looked at that there is a correlation between those who have higher rates of suppression and having greater retention. We have seen, though, I will be honest with you, we have seen, we used to see significant gains on our retention from year to year and I think you can see that as evidence for some of our data and we have seen retention rate sort of plateau.

But I want to remind is that, we're using the Ryan White HIV/AIDS Program data as well as the (Hebron) data. And the (Hebron) data are collected primarily as sites that are funded through the Ryan White program. And we know in the Ryan White Program, we have much better performance outcomes and performance data compared to the general HIV population.

And I kind of would go back to, when we look at the care continuum data that's presented by CDC. And I think as Kathleen had commented earlier. Although, when we're looking at the Ryan White data, we might not see significant gap in care. When we look at the CDC data, we are seeing significant gap and that 86 percent of people who are diagnosed. And then, we see a following off in terms of 40 percent of people are retained in other drop off for prescription of antiretroviral therapy and viral suppression as well.

Adam Thompson: Thank you.

Nanette Benbow: I have a question. The questions that are being posed now, I had many of them. I addressed them in the last section that has to do with use. Should I just hold off and discuss it there or talk about it now?

Christy Skipper: So we'll go in order of the criteria, so it sounds like we're wrapping up discussion on scientific acceptability and we can move on to feasibility and then usability and use.

Nanette Benbow: OK, sounds good.

Christy Skipper: Thank you.

Kathleen Brady: So, I guess, in terms of the summary for feasibility, in terms of the specifications including logic require data that are readily available and can be easily implemented without undo burden. And I may not like found that through locally, and it – I agree with the preliminary rating for feasibility that it is indeed high.

I don't know if you have any questions or comments regarding that, Nanette.

Nanette Benbow: No. No, I agree.

Kathleen Brady: And then onto usability and use, so basically, this indicator is being currently publicly reported and current use in accountability program both including the quality reporting system and value-based modifier and merit-based incentive payment system. The information within the HIV (RM) shows that the data has improved over time as we have indicated. And that – since the development of this measure, it has been actually adopted by CMF measurement programs and selected as a core HIV indicator by the secretary of the Department of Health and Human Services. And the National Learning Collaborative through HIV quality improvement activities has also used the measure for Ryan White's HIV/AIDS Program grant recipients and separate recipients.

I agree that I don't know that there's necessarily any harms. And they are getting feedback from Ryan White's HIV/AIDS Program recipients regarding the indicator. And I think that's it. And I agree with the preliminary ratings for usability and uses high.

Christy Skipper: And it sounds like – was there question about usability from the committees?

Nanette Benbow: Yes. So – And I agree. It sounds like the developer obtained feedback. You know, from multiple parties including Ryan White HIV/AIDS program recipients who thought it was a helpful performance measure.

A couple of things that I found interesting was that – so while it sounds to be used in a number setting, I don't see the use of this particular measure in the literature. What I see more is and sort of what's been quoted in previous conversations this morning is that when we talk about the relationship at the

same retention and viral suppression, those examples have to do with why CDC is defining the measure and not the way that this measure is defined.

So there are different timeframes. One is over across the two year of span. The other is a one year span. And so, you know, if one wanted to use all the measures that are being discussed today, it seems that prescription of antiretroviral and viral suppression are within a measure – or a year, a measure – the measurement is one year where this one is two years.

So I know there's a question but harmonization and I apologize, I'm new. I don't know if this is exactly fits the criteria of harmonization. But it does seem that the fact that this measure is over a two-year period is odd. So I just wonder how feasible it is that Ryan White, you know, program recipients would use this when they're looking at all the other measures within one year.

And kind of I think related to that is that the other measures do use Ryan White services report, RSR data, to talk about their measures and discuss their performance. But this measure does not use RSR data which, you know, as (Marlin) said, it's an incredibly rich and important way of being able to monitor what's happening with Ryan White Program recipients.

So I guess as I'm looking at all of this, I wonder how much it is actually usable on the ground.

(Marlin Matoski): May I respond?

Christy Skipper: Yes, you may.

(Marlin Matoski): Thank you. So, Nanette, I think I'm going to go backwards. I'm going to take your – I felt like there were two questions there.

Nanette Benbow: Yes.

(Marlin Matoski): And I'm going to – I felt like the first one was, "Why didn't you use RSR data when you were testing this measure?" I feel like that was the latter question. And then the former was, "Why is it 24 months?"

I'm going to start with the RSR question because that's a little bit, I think, much more straightforward. So for folks who are aware or those who aren't, the Ryan White HIV/AIDS program collects a Ryan White services report annually from all funded grant recipients and providers. We have any – We have around 600 to 700 grant recipients annually which translates into over a thousand providers annually. When we collect those data, we're looking at services and related client data for a calendar year. It is our full intent that we will be able to link clients across reporting period but at this time we're not able to.

So hence, we're able to, you know, for the most part, calculates performance data based on our RSR or our Ryan White services report data for any measure that's looking at a 12-month reporting period because that's how we are currently collecting the data. But we're limited at this time, looking at measures that either look back into the previous reporting period to populate a denominator or what have you, so that's number one.

We hope to do that in the future. We currently are actively – we've been actively working on it for, you know, about a year now. And we hoped to have a data solution to this problem.

We've used the HIV research for this particular measure because are, in fact, able to look at clients over a great length of time. As many of you know, the research network has been around for a very long time and they have, you know, a nearly a decade of data, longitudinal data on clients. And so, they're ably – that are easily able to look at clients and their related data over a 24-month period of time. So that's the answer to that question about why didn't we use the RSR data.

Here are the question about why is this different, why is this measure different in terms of the reporting period, i.e., all of the other measures are looking at a 24-month measure, or sorry, 12-month measurement period and this one's looking at a 12-month, or sorry, let me back up. This one's looking at a 24-month measurements period whereas, most other measures look at a 12-month.

You know, researchers indicated that retention isn't something that you just do for one year at a time and specifically within HIV. HIV is considered a chronic illness that folks live with for many decades at this point. And so, it is our intent with measure to start longer term retention. We do know that there are patients who are coming in and out of care for a variety of reasons. And that coming in and out of care can have an impact on their health outcomes, and more specifically their morbidity and their mortality.

So this was our attempt at looking at more longer term retention. And I used longer term in air quotes. There is another performance measure that exists in the general performance measure realm. It was related to cervical cancer screening that had a longer look back period. And that measure is being used by CMF or the Centers from Medicare and Medicaid in a variety of programs.

Nanette Benbow: Thank you, that makes sense. I think just as someone who was work to the Health Department and, you know, wanted to look at all these different performance measures. I think I am not alone in saying that Health Departments encounter – and Health Department to fund, you know, Ryan White sub-grantees locally, experience the same problem that it sounds you are having, (Marlin), of combining data across the years. So it sort of the rationale makes sense, but it's – it can be logistically difficult to calculate.

Now, I understand CAREWare has, you know, can provide that measure as part of it summary data. But, you know, I don't know how difficult this might be for those that who don't have CAREWare.

Adam Thompson: Yes. This is Adam. I have a follow-up question for that and it also had to do with the two year frequency and it also has to do with what they expected use of the clinic level is. I know, I've seen a lot of folks in the Ryan White system and disclosure I work in the Ryan White system have a lot of challenges actually using the data to inform their process. I've seen more people kind of moving towards the gap measure versus this medical visit frequency measure.

And one of the issues is that, a lot of the clinics that I'm working with, they seem to have identified that when an individual is lost to care that immediate

response to a med-visit is more actionable and more meaningful than waiting for a retention measure that shows it over time.

Also, when they are looking at moving a patient off their active list, so that they are no longer a patient that that process seems to be happening at the year mark instead of a two-year mark, and I'm concern that if this becomes part of MIPS and the MACRA incentive payment that individuals may be held accountable for performance on patients that are no longer there.

(Marlin Matoski): That was a very interesting points that you made there, Adam. I want to reiterate part of what you said, we do see that these measures operate in more of a suite of setting, meaning that they would operate and be used together. As you would indicated the gap and visits measure is looking at a shorter duration of time and it's able to give you much more actionable data.

But we also, you know, as I've said earlier we didn't want folks to lose sight of that longer term retention. Because we did know that some folks were mostly focusing on the shorter term retention and they were loosing people in the long run. And we've done, you know, just exclusively in the Ryan White Program, we've worked with our grant recipients on how you can operationalize both of these measures and how they would function and form your program in different ways.

Adam Thompson: Thank you.

Christy Skipper: All right. So it sounds like we're having a really good discussion about this measure. And it sounds like we're ready to wrap up and move on to the eMeasure version of this measure 3209. And you already had the discussion about the evidence and the performance gap on the paper-based measures, so your comments and thoughts there also applied to the eMeasure. So we'll start out 3209 in regard to the eMeasure specification.

Kathleen Brady: OK. So that – in terms of the specification that under reliability.

Christy Skipper: Under – yes, under reliability.

Kathleen Brady: Nanette, do you want to start with this?

Nanette Benbow: Oh, go ahead. You're doing such a great job, Kathleen.

Kathleen Brady: OK.

Nanette Benbow: Thank you.

Kathleen Brady: OK. I'm less familiar with this one than I was with the first one, OK. So in terms of the specification, the eMeasure specifications were included, you know, the same submission and basically that the …

(Off-Mic)

Kathleen Brady: … did follow the industry accepted format for eMeasure in HL7 Health Quality Measure Format, the HQMF. So, and the submitted eMeasure specifications used existing value sets when possible and use new values that have been vetted through the (VSAC). Submission includes test results from a stimulated data set, demonstrated the measure logic, it can be interpreted precisely and unambiguously. And they submitted feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on the assessment by EHR vendors.

So a lot of that I didn't understand, no, I shouldn't say that but, it can be done is I think the answer to that one. And EHR people said yes, it can be done. So that's really what the most important thing

In terms of reliability testing, the testing level was at the data elements. And so it was perform – a data set including 64 synthetic patients created in the Bonnie testing system – which I stated in the Bonnie system – hearing an echo …

Christy Skipper: Excuse me, if you can press star six to mute your line.

(Off-Mic)

Christy Skipper: OK. I believe it's been corrected. So we were on the reliability testing with the 64 patients.

Kathleen Brady: So yes. So the use of 64 synthetic patients and they tested certain data elements in the Bonnie testing tools not – I think one of the recommended tools for doing the reliability testing. So they used name, data of birth rates, ethnicity, gender payer diagnosis and encounters. And the patients bundle – patients bundle demographics were design to mimic the HIV/AIDS population and were derived from RSR type data. And data element, validity testing was performed and transfer data element reliability.

And basically, the reliability results from the paper-based version of this measure and currently there are no performance data available to test the ECQM. However, the chart extracted version of this measure has been in use in a National Quality Reporting Program since as early as 2010.

Melissa Marinelarena: Hi. And this is Melissa. I just wanted to remind everyone that the testing requirements for this, what we consider legacy measures which are electronically like re-specified electronic critical measures based on existing measures that are in federal program. We don't require the type of testing that we do for a newly specified eMeasure. So these synthetic testing is sufficient which test really the logic of the measure and feasibility.

And like you mentioned, the Bonnie testing is sufficient. So it shows that the measure – that the measure logic does work. It is able to calculate in a synthetic patient test bed. I just want to remind everybody of what the difference is there and we'll also do another presentation during the meeting to discus that.

Kathleen Brady: OK. And so – and moderate is the highest rating possible for that.

Melissa Marinelarena: Correct, correct. Because they can't really do through reliability or validity testing, you know, no measure score because they were – no actually, you know, facilities that they are doing. They are in the process of testing these measures.

And then for data elements, you know, this is close so that they could do, they do – they did do a very good job of mimicking this patient, this synthetic

patient-based on true patient that exist in – within Ryan White which was great. But, you know, through data elements testing or through validity testing isn't possible, but this is sufficient for a legacy measure.

Woody Eisenberg: This is Woody. I have a question for the technical developers. How did you choose the number if synthetic patients in the 60s? Was that to represent many different combinations of patient characteristics – and some redundancies?

(Marlin Matoski): I'm going to ask (Ruth) from our team. She was one of the leads on the ECQM work, (Ruth), could you respond?

(Ruth Martin): Sure, absolutely. So what we try to do with the body testing and with the synthetic patients is to cover the entire logic. So that for each inclusion and exclusion criterion, we have at least one patient that will meet the criterion or not meet – and not meet the criterion. And then we also include patients that have – that are close to the timing thresholds of each of the inclusion and exclusion criteria.

So for instance, that three months mark for the diagnosis, we try to create test patients that are around those. And these test patients are really designed to make sure that we, in logic, we've design to be computed automatically matches the intent of the original measure. And so, the way this work is we define the expectation for each patient. And then, we run the patient data through the measure logic and see if it matches the expectation. And we did get a 100 percent match and we did get a hundred percent coverage in the logic as I've stated.

And then in addition to that, we also include additional synthetic patient. Actually all of the sample had – we strive to mimic the characteristics of the Ryan White population. But then we also added a few more realistic patient scenarios that weren't so driven at testing the logic in individual portion but as an overall clinical scenario. And so, it's a mix of both vary patient characteristics and then scenarios that will help us – that will enable us to test the full breadth of the CQM logic if that makes sense.

Woody Eisenberg: Thank you.

Christy Skipper: OK, moving right along to feasibility. Are there any questions, comments about feasibility?

Woody Eisenberg: This is Woody. Do we know how many different kinds of electronic health records would support this measure, beyond Ryan White's in particular?

(Ruth Martin): So like – this is (Ruth), I can speak to that as well. So, what we did for the feasibility score card for the measure right now is just the panel evaluation. So that's how we filled out the score card. And we include both ECQM, Electronic Clinical Quality Measure experts, as well as health I.T. experts, nurse informaticists that have worked in the field extensively and sort of natural thing. And try to rate the data elements and how they rated it in the criteria for the score card that way.

We are currently working on – excuse me. We're currently working on actually engaging directly with each of our vendors and with clinical sites. And we're looking at engaging at least with three different clinical sites and three different EHR vendors. And hopefully that will give us at least six EHR vendors coverage which coincidentally covers, I think, about 60 percent to 75 percent of providers just based on meaningful used data. So, what we're trying to do is make sure that we're going to the EHR vendors that are represented of the market shares in U.S these days.

But for the purpose of the endorsement right now, we did use a lot of the meaningful use data that was available in terms of both the functionality that is required for meaningful use. And also the measures that are currently part of the meaningful use program. And I think all except for one data element are data element that are used in other clinical quality measures that are currently part of the meaningful use program. So, that's probably more information that what you were looking for.

Woody Eisenberg: No, that's just what I was looking for. Thank you.

Nanette Benbow: Yes. I appreciate that as well. This is Nanette, that I have the same question. And I think you've answered it perfectly. Thank you.

Christy Skipper: I think the last criteria for this measure usability and use. And just looking at some other pre-evaluation comments that you all committed I see that one comment says that you all had similar comments on usability and use for this measure as you did with the paper measure 2079, unless Nanette or Kathleen, committee would like to add anything to the discussion that we'll move on to 2080.

Kathleen Brady: I don't think so.

Nanette Benbow: Yes, no, I'm fine.

Christy Skipper: OK. Thank you. So we just finish the first set of measures and now we'll move on to measure 2080 Gap in Medical Visits. And our lead discussants, again, are Kathleen Brady and Jeffrey Lewis. So I'll turn to either one of you to start off by introducing the measure and telling us what it is and following the same format that we just went through. So take it away.

Kathleen Brady: Is Jeffrey on the call? Hello?

Nanette Benbow: No, Kathleen, it's all you.

Kathleen Brady: OK. Oh, yeah.

Nanette Benbow: Now, I feel really guilty, Kathleen.

Kathleen Brady: OK. So, well, this is actually a measure that I use pretty much on a daily basis, though I know a lot about this one. So, anyway, so 2080 is Gap in HIV Medical Visits, HRSA again is the measure steward. And the brief description is the percentage of patients regardless of age with the diagnosis of HIV who did not have medical visit in the last six months of the measurement year. And so, it's basically – it is people who had a visit actually in the first half of the measurement here that didn't have a visit in the second half of the year. So, you can look for gaps in the care.

And so, in terms of the rationale, they provide some of the same information that was provided in the previous regarding retention playing a critical role in folks achieving viral suppression. And we're just trying to see what else we can measure here.

And I think that's really about it, you know, and they do highlight the information here about in terms of where transmissions come from along the HIV care continuum. And the study by Jacek Skarbinski that actually showed that 91 percent of new infections in 2009 were attributable to people who actually were infected with HIV and unaware which is about 30 percent. And I think it was 61 percent of the infections actually were in people who are HIV infected aware of their status but not engage to medical care.

And very few infections are actually come from folks who are engage in care on ART and virally – very few. We know (not) pretty much, none from folks that are virally suppressed. And so basically and they highlight that list of patients and the numerator can be generated a medical provider staff can immediately begin follow-up with the patient to schedule an appointment for a medical visit. So, this is once again a profit measure.

Christy Skipper: OK. And this is also a maintenance measure. And so in the discussion of evidence, if there is no new evidence in the committee is comfortable with what submitted last time. We can talk about it but at the end person meeting, you don't necessarily need to revote on it. And then, again the emphasis with the maintenance, the new maintenance process is on the performance gap. So, if you all agree or if there's no anything to add for the evidence we can go ahead and talk about performance gap for this measure.

Kathleen Brady: I don't have anything to add.

Christy Skipper: Is that for performance gap or just the evidence?

Kathleen Brady: No, I don't have anything to add in terms of evidence.

Christy Skipper: OK.

Melissa Marinelarena: And this is Melissa, just a quick note. Usually the way we operationalize it in the medians is, if at least one person wants to vote, revote on the evidence, we will hold the revote just to be inclusive. But if everyone agrees not to revote then that usually how, then we don't have to revote. But just as note, and we'll talk about it again at the meeting but that usually how we operationalize this process.

Christy Skipper: OK. Kathleen, sorry to interrupt, go ahead.

Kathleen Brady: That's OK, all right. So, in terms of performance gap, the developer presented data from the RSR and which we we've already heard about. And the mean performance for gaps in medical visits has fluctuated over time but stands at 21.7 percent as of calendar year 2014. They show the number of patients and number of providers over time. Number of providers has decline a little bit, but the number of – actually, the number of patients declined in 2014. But the median has actually been pretty similar over time, has slight increase in 2014, but I'm not sure if that's the statistically significant. And they show basically in terms of gaps, disparities and medical visits among person age 20 to 34 among Native Hawaiian Pacific Islanders and American Indian Alaskan Natives and transgender patients. With the highest gaps, you know, really among those populations.

Christy Skipper: All right. Any questions for the committee – from the committee about performance gap? All right, we can move on to reliability.

Kathleen Brady: OK. And so, this – for this measure, it's for paper records. It's not the eMeasure. The measure is testified at the level and a clinician office or clinic. And so, patients are included in a numerator if they don't have a medical visit in the last six months on the measurement year, and then denominator includes number of patient who had at least one medical visit in the first six months of the measurement year. Patients are excluded if they died and the measure calculates a rate where a lower score is associated with better performance.

In terms of reliability testing, basically, it's the testing level is at the measure score. Measure, yes, measure score, I'm sorry. And the developer estimated

reliability using a beta-binomial model and basically, the testing result showed a median reliability of 0.973 in 2014 which is very good reliability. And they also gave a distribution of provider level reliability scores by year, which were good not quite as good as the previous measure.

But the overall it was a rating of high which I agree with. Do you want me to continue?

Christy Skipper: Yes.

Kathleen Brady: OK. In terms of validity, in terms of the testing, the testing is based on face validity only. And it's not different I think than what was submitted previously. OK.

Christy Skipper: OK. Did committee members have any questions or comments about reliability or validity, if not we can move on.

OK, moving on to feasibility.

Kathleen Brady: So, the developer reports that all data are generated and collected by and use by health care personnel during the provision of care. Developer reports that all data elements are in defied field and electronic health records and those fees licensing or other requirements to use this measure.

So, it's pretty feasible. We've been actually using it at a low, at the local level not just using it from our CAREWare data but actually into data to care projects at the Philadelphia Department of Public House and it's very easy to implement.

Christy Skipper: OK. And just look – taking and look at the committee pre-evaluation comments, I do see a comment and there I don't know if the individual who submitted this comment or question is on the phone. But I'll read it, it just says, this is a paper-derived measure, it's hard to know how complete patient appointment data is how are missed or canceled appointments is counted. So, if you entered this comment and you still like to post this question or you want to provide a little more detail, please feel free to do so.

Woody Eisenberg: Yes. This is Woody Eisenberg. I think that was question. And it had to do with what happens to canceled appointments, do they just drop off the map? How do you include those?

(Marlin Matoski): So, for this measure, it's a measurement that actually patients who actually attended visit. And so if in the instance that a client missed the visit or if a visit was either cancel and postponed or postponed or what have you, then that would not be captured because all we're looking at are attended visits.

Woody Eisenberg: So, whether a patient just disappeared off the map or scheduled an appointment and didn't show, or scheduled an appointment and canceled, it would all be picked up simply as not having attended a visit in the last six months.

Christy Skipper: Correct.

Woody Eisenberg: OK, thank you.

Christy Skipper: Any other questions? OK, I guess we can move on to usability and use.

Kathleen Brady: So, the measure is currently being publicly reported and it is not in use in an accountability program. And so, it is used in the Ryan White HIV/AIDS program and – so the – they do point out that as retention and care has may be been less of our priority that the Ryan White HIV/AIDS Program has expand at three point increase in gap in medical visits from 18.6 in 2010 to 21.7 percent in 2014.

And so, the measure was selected as one of the core HIV indicators by the secretary of the Department of Health and Human Services. There are no potential harms that have been identified and once again they have – they have gotten feedback from the Ryan White HIV/AIDS program grant recipient in terms of vetting the measure.

And I think that is it in terms of feasibility – I'm sorry usability and use.

Christy Skipper: OK, all right. So we will just continue to scroll down the worksheet. I just want to point out the pre-meeting public and member comment. So, in

addition to your comments that you all gave us, we also open the measures to members in the public and we did receive one comment on this measure.

And so, you can see it there in that grayish colored box gaps in care should be focus on minimum standards applying to all patients as absence of a clinical visit within a 12-month period. So, and just some – this is the section on the worksheet where you can see and respond to meeting or comments that came in during that commenting period.

Kathleen Brady: Yes, I mean, I can tell you my thoughts on that and that is the longer someone who actually has been out of care that should have been in care, the less likely it is that yore going to be able to find that individual. And so, I don't agree with that. I think the six, you know, six months gap is appropriate.

You know, I can tell you from our local experience we used it in conjunction with viral suppression. Actually, mainly looking in durable viral suppression and also follow-up with providers to see if there's been any other ongoing, you know, issues, as there are other comorbidities that would require the patient to have been in care, you know, more often, or if there a well patient excluding them from follow-up and allowing them time to come back to care. And we find very effective.

Adam Thompson: Yes. This is Adam. I would echo similarly. I think what confound me about this is, is when we think about them as a suite with retention measure. Because I'd see it as very useful, you know, identifying individuals who are in need of care and as we move from reactive to proactive thinking, this is a really good way to proactively identify individuals, you should reaching out to schedule visits instead of waiting to see them further lost in care.

But, I think it also sets you up to meet the other measure as well. So, if we accept the retention measure, as a good measure of retention, I think this measure sets you up to kind of get there in the process. But, I think it also – it's important to just consider the possibility that some folks are only going to be seen once a year. And really how these measures relate into that performance payment program to make sure that it's not so many people that is going to affect the provider's ability to risk those incentives.

Kathleen Brady: Yes. What we always tell our providers at the local level that we never expect the gap measure or the retention – well, the gap measure to be zero and then retention measure to be 100 percent. And so, you know, that's what we're expecting from a quality management perspective as the grantee, Ryan White Part A grantee, for the Philadelphia EMA. But, I think that in terms – if this ends up being something that's related to reimbursement through CMS, I think, you know, and same for the retention measure that, you know, the bar for reimbursement or increase incentives has to be appropriate.

Adam Thompson: Yes. The other comment that I would add too, and I agree with everything Dr. Brady said. The other comment is around the fact that this is an inverse measure. And I get it, I understand it and I don't really have an issue kind of looking at it and for the lower number being the better performance, but I have definitely heard from the field lots of people have a problem interpreting the result of this and they spend a lot of time explaining to folks. So, I just want to kind of put that out there in the discussion that this does somewhat confound people no matter how simple it may seem to us.

Kathleen Brady: Yes. That is very true. We have problems with our providers on a local level that, you know, are – start looking, you know, with the other list you start looking at what we call the not enumerator list too. And so they're trying to do that with this list, but it's actually the enumerator list that you really want to evaluate. And so they end up spending some resources, finding out that, yes, those people are actually in there.

Adam Thompson: Yes.

Melissa Marinelarena: Hi, this is Melissa from NQF. I also want to remind everybody that NQF does not endorse list of measures. These measures have to stand on their own. You can make recommendations that they are grouped together or paired together for reporting, but we do not endorse them. We would not endorse them as a suite of measure. So I just wanted to make that clear.

Adam Thompson: Cool. Thank you.

Melissa Marinelarena: Thank you.

Christy Skipper: All right. So, now we'll move on to measure 2082, HIV Viral Load Suppression and our lead discussants are Laura Evans, Piero Garzaro and Michael Lane. And I'll leave it to you all to decide who would like to kick off the discussion.

Male: Who want to start?

(Crosstalk)

Laura Evans: This is Laura. I can start unless Kathleen wants to do this one, too.

Kathleen Brady: No, I'm done.

Laura Evans: OK. Well, I will call liberally on my colleagues and peers here to help me out here. But – so this is measure 2082. The measure title is HIV Viral Load Suppression. Again, the steward is HRSA and this is a maintenance measure.

The brief description is this is a percentage of patients of any age with a diagnosis of HIV who have an HIV viral load less than 200 copies per ml at the last viral load test during the measurement year.

The rationale behind this and there's actually in the measure worksheet a logic model provided that basically outline steps between – of medical care that people living with HIV go through from diagnosis to achieving a goal of viral suppression including diagnosis linkage to care, retention care, receiving antiretroviral therapy and then viral suppression and that notes these models that incorporated into National HIV/AIDS Strategy and is focus all HIV prevention care and treatment efforts in the U.S.

So the rationale is that, suppression of viral load, I think as we've talked about with some of the other measures leads to improve outcomes for patients with HIV, and so – in contrast to the previous measures that we've discussed, this is an outcome measure. Anything you guys want to add in terms of the basic measure description?

Male: None. That summarizes pretty well.

Male: Yes, I agree.

Laura Evans: OK. So then, do you want to go on to evidence or is there any other question from the group?

Christy Skipper: Correct. We can move on to evidence. Again – yes.

Laura Evans: Again, this a maintenance measure around that, so there is some updated evidence that was provided by the measure developers, which is predominantly in the form of updated guidelines. The previous rating of the evidence was sufficient demonstrating that antiretroviral therapy and viral suppression reduced morbidity and mortality associate with HIV.

And so the preliminary rating from this was, this is a past, the updated guidelines again link viral suppression to improve outcomes there. I do – did note that in the pre-evaluation comments from the committee, there were a couple notes there about the evidence, so I know that we are emphasizing this a little bit less, but I thought maybe they – (there's) a little bit of discussion.

I see two comment related to querying about the viral load threshold being setup less than 200 copies per ml now in the setting of more sensitive testing of viral load and whether that is still the appropriate threshold or whether the threshold should be decrease to undetectable or, you know, lower numerical target of less than 50 or less than 20 copies per ml.

Kathleen Brady: No.

Male: Yes, I would agree.

Kathleen Brady: No. You know, it's set at 200 weeks so that way you don't inadvertently call someone detectable who had a blip and their viral load is 20 or 21, which I think most of us sort of agree is not actually and it's – the test isn't that good that he can tell the difference between less than 20 in a viral load of something like 21.

So, there are some people who still use (BDNAs), where the threshold is less than 75 and we actually know, you know, the transition threshold is actually about 1,500. So, I think less than 200, you know, for clinical purpose is, is appropriate.

Melissa Marinelarena: Hi. And this is Melissa from NQF. And just a reminder, this is an outcome measure, so the requirement for the evidence is a little bit different. They just have to provide the rationale. So this is either, yes, they have, or no they haven't, so pass, do not pass. You know, that is a great question and I know I look for that and there is information in the guidelines, but it's just whether you agree or do not agree with the rationale, though it's a little bit different on the evidence for an outcome measure.

Laura Evans: Thanks for that clarification. Shall we move on to the gap then, the performance gap?

Christy Skipper: Yes.

Laura Evans: Some other queries? OK. So, there was, again, data from the Ryan White Service Report, the RSR, that we've seen from some of the other performance measures as well, again, over the same time – five-year time period from 2010 to 2014 and there's a table provided here with more recent data from 2014 on the left and going across to the right going year by year across that.

And basically, reports the mean and median patients for meeting this measure of viral load suppression of less than 200 copies per ml and you can see that there's improvement over time. So on the more – of recent counts, 2014, the proportion of patient achieving viral load suppression is higher in each sequential year going back to 2010, although there's still a significant gap there was about – with a mean of 80 and a median of 84 percent of patients achieving viral load suppression.

And if you go down the table to the next table, it breaks down some of the disparities by age, race, ethnicity and gender as well. And you can see that there are some significant differences in achieving viral load suppression by all of this categories of age, race, ethnicity, sorry, and gender as well.

So this (inaudible) the previous – preliminary rating for opportunity for improvement of high and I personally agree with that. I don't know if my other colleagues had anything they want to weigh in on about the reporting of the performance gap.

Male: I completely agree.

Male: Yes.

Laura Evans: OK. Any other comments from the group? OK. And, Christy, should we move on to reliability there?

Christy Skipper: Yes.

Laura Evans: OK. So if you go down to reliability, this is, I think somewhat similar to the previous discussions that we've had. This analysis is at the facility level. The numerator is the number of patients in the – with the viral load less than 200 copies per ml at the last test or any measurement year. The denominator is the number of patients regardless of age with the diagnosis of HIV with at least one medical visit in the measurement year. And to be included in the denominator, they have to meet the following criteria any age, diagnosed with HIV during the first three months of the measurement year or prior to the measurement year. And have at least one medical visit during the measurement year. There are no patient's exclusions.

Piero Garzaro: This is Piero. I have a question, this is for the saying for the group. When you say at least one medical visit, do you mean in person or any type, I mean, now in technology age we have video visit, we have telephone appointment visit, what do you mean by a medical visit?

Christy Skipper: (Marlin), would you like to answer that?

(Marlin Matoski): Sure. So we are capturing a medical visit as a – of a reaction between a provider or prescribing privileges and the client. For the chart abstracted measure, I think it's been mostly safe to say a visit in the clinic setting.

However, as structured data is able to capture telemedicine visit, we will be considering adding those into the ECQM version of the measure.

Piero Garzaro: OK.

(Marlin Matoski): (Ruth), is there anything you would add about that in terms of the ECQM in capturing telehealth visits or none face to face visit?

(Ruth Martin): No, not at this point …

(Off-Mic)

Male: No.

(Ruth Martin): … except there.

(Marlin Matoski): And I just want to know – I know we're going to talk about ECQM next, but I'll bring it up, because it came up here. And (Ruth) probably can explain this much better, but the value set for our visits is applicable not just HIV measures that's applicable to many other ECQMs. So we don't own that value set. So, many, many measure developers use that medical visit value sets so we're somewhat at the mercy of the owner of value set.

Although I say we're at the mercy of that owner, it does lead to greater alignment of measures in general in terms of we're all looking at the same "definition of a medical visit".

(Ruth), is there anything you would add there?

(Ruth Martin): No, I think you explain it beautifully. The value sets for encounters are shared across most if not all the meaningful use or the now defunct to meaningful use measures. So we did strive to align as much as possible. As I mentioned earlier, you think all that one value set use in this measure is already use in other measure. So, we're striving for lowering the bar of implementation and the confusion that often comes of slight misalignment across measures and oftentimes similar measures. So we've definitely prioritize that. And as of

right now, it's just checking that the measures include face to face visits. So the value sets cover essentially face to face visit at this point.

In terms of telemedicine and none face to face visits, I think that's definitely an issue that has come up across multiple measures. So I definitely recommend that this is look at in a broader sense rather than trying to address it measure by measure, so that we can maintain that alignment on the electronic measurement site for sure.

Christy Skipper: OK. It sounds like we can move on to validity.

Male: Can I just ask one follow-up on the specifications for the numerator and denominator statements. Just to clarify, there was an internal inconsistency and the statements in the brief overview denominator description doesn't include any of those conditions that are stated in the full statement. And the measures, this measure that charted abstracted on the electronic measure don't match in those sections.

I think what (inaudible) through the rest of the packet is that all of those conditions apply the patients, you know, any age but specifically diagnosed within the first three years, three months in the measurement year or the prior year included. That statement isn't included, made me take a little bit of a double take when I read it in the brief description above and then the cross comparison that statement is included in the eMeasure. Just perhaps before the in-person meeting, we can address that to make sure that the statements are clear.

Christy Skipper: OK, so are you pointing out what I've highlighted on the screen? Does not match what is above in the measure denominator statement is that what you're pointing out?

Male: No, on my screen, I am still seeing …

Christy Skipper: Reliability?

Male: … the prior – the prior measure. So.

Christy Skipper: Hit refresh at the top of your screen.

Melissa Marinelarena: This is Melissa. It might be because of the way the wording is for eMeasures. It calculates like the inpatient population is included in the numerator, I'll take a look at it and make sure that it's consistent but he did a really good job of making sure that they were consistent but it's just they are measuring the same thing, it's just the way an eMeasure calculates the population and then the denominator and then the numerator. But I'll take a look at it and make sure that that we did capture all the records.

Male: Yes. So the languages consistent across both measures, but the brief description for the chart abstracted denominator statement doesn't include the – those conditions whereas the eMeasure and that same section does.

Melissa Marinelarena: OK, I'll take a look at it to make sure that …

Male: Yes.

Melissa Marinelarena: … that we got right, thank you.

Male: My OCD kicking in.

Melissa Marinelarena: Oh, I have it.

Laura Evans:: I have really good eyes. So, in terms of reliability testing, the methods of reliability testing were similar to the measure 2079, now we saw at the measure score based using the RSR data set. And they report again using update, I know I mean a model to assess the signal to a noise ratio and report very good reliability of the testing over time. And again the preliminary rating for reliability was high.

Christy Skipper: OK, any questions, comments from the committee? All right, we'll move on to validity.

Laura Evans: So the validity is also similar to what we saw with the previous paper measures where this is face validity testing of the measure score, the bases of the technical workgroup and panel for the development of the measure.

And again, I think, I mean I don't know if (Marlin) and that the same emphasis applies here with moving towards the eMeasure as opposed to doing numerical or, you know, empirical validity testing on the paper measure that.

(Marlin Matoski): Yes. So we've been – we've sort of pivoted, we had to continue to maintain these measures, obviously because they had been currently endorsed and we wouldn't have been able to brought the – bring the eMeasures for endorsement if we didn't have existing legacy measures. So, we continue to do the very basic routine maintenance on the – I called them the paper measure or the chart abstraction measure. But we've pivoted our resources into developing and testing all of the ECQM.

So we have an ECQM that's being field tested for each of these measures that are up for maintenance.

Christy Skipper: Any other questions from co-discussant, comments from the committee?

Adam Thompson: I have one question under the face validity, and this is more – maybe just a process or clarification. Under the risk adjustment section, it says they're not risk-adjusted and they gave justification based on the demographic served by the Ryan White population which makes sense. This is just around consistency in statement.

They say at the end additionally this measure is not use for paper performance bonuses or penalties. But then later they indicate that if contention to be use in MIPS program. So I just wondered whether those two statements agree or whether one is inconsistent with the other.

Christy Skipper: I think …

(Marlin Matoski): Oh sorry, go on.

Christy Skipper: So, go ahead (Marlin), go ahead.

(Marlin Matoski): So, thanks Adam for that comment. So at this point, we have not risk-adjusted any of our measures that you had suggested and there are no number

of measures that are in various center for Medicare and Medicaid programs that are not risk-adjusted.

We feel as though once we've established the baseline performance, we will be able to work with CMS to likewise set baselines for performance and not proceed with any risk adjustment. Because, you know, quite honestly, if we are risk adjusting for some of the very classically risk-adjusted categories or patient characteristics, we may (be) doing …

Adam Thompson: The whole population.

(Marlin Matoski): Yes. You know, that's the thing, you know. And so, you know – and so what really would do with those sorts of things. And, you know, I think that, you know, without, you know, giving, you know, a master class on this, you know, HIV does disproportionately affect many of the populations are risk-adjusted for. And so, we've had very often and very philosophical discussion about this internally as well as with some external colleagues.

Adam Thompson: Thank you.

Christy Skipper: All right, moving on to feasibility.

Laura Evans: OK. In the feasibility, we have report that the data elements are all generated during the routine provision of care and are available in the site – fields in the electronic health record. The data are readily available within patient health records and provided annually to the Ryan White Program through the RSR report and no fees, licensing or other requirements to use. And the preliminary rating was high for feasibility and looks like the preliminary comments feel similarly that this – where it's a high rating for feasibility.

Christy Skipper: All right, if there are no comments or questions, we'll move on to usability and use.

Laura Evans: Finally, usability and use, there are current uses of the measure are publicly reporting and it's currently in use in the accountability program and there are several programs listed here, including the Ryan White Program, the Medicaid

Adult Core Set, PQRS and Value-Based Modifier Program, the MIPS Program and the National HIV/AIDS Strategy Program.

There's a note made that the Ryan White Program has experienced the 20-point increase in viral suppression over five years from 2010 to 2014 with increases across all demographic groups and some populations with no unexpected findings during implementation. And no statement that potential harms were identified during implementation with that.

So, I think the general feeling certainly from the preliminary rating and from my reading of this as well is that preliminary rating for usability and use of high seems appropriate given that extent of adoption into five different accountability programs.

Christy Skipper: OK. There were no pre-meeting or member commenting – comments received on this measure, so we can jump right into the eMeasure discussing the measure specifications for 3210 Viral Load Suppression, because you've already had the discussion on evidence and gaps for the paper-based version of this measure.

Laura Evans: Piero or Mike, do you want to take this one?

Michael Lane: You're doing so well.

Laura Evans: Come on. Help the person out. All right, we'll muddle through it together.

Piero Garzaro: I'll jump in a second. I'm just pulling it up.

Melissa Marinelarena: We'll put them – this is Melissa. We'll put them on the spot during the in-person meeting. How about that?

Michael Lane: OK.

Laura Evans: Deal. OK. So, this is – 3210 is, again, the sort of eMeasure accompaniment to the measure we just reviewed of 2082. So it's the HIV Viral Load Suppression, again the measure steward is HRSA. And the measure – do we

want to go through the rationale and everything again because I think it's basically the same?

Christy Skipper: Right, it's the same.

Piero Garzaro: Identical.

Christy Skipper: Yes, it's identical. We want to go to the reliability specifications.

Laura Evans: OK. So let's jump right to reliability with this. And so, I feel like Kathleen and her previous discussion here, so I'm going to kind of read what's on here but I'm not sure and I'm not super familiar with.

Male: Let me – before you start, I mean, what – I think that many of us when we saw that it was the same, saying – we said, "Oh, they're identical." At least, I couldn't understand why they have two things that look essentially the same. So what – how an eMeasure – what's the basic different between an eMeasure and regular measure? I mean, they measured the same thing, their data is the same, so what's the score difference between an eMeasure, how can they differ?

Christy Skipper: (Marlin), do you want have (Ruth) to explain this or do you want the explanation from the NQF perspective.

(Marlin Matoski): I would prefer if (Ruth) because (Ruth) has a really – from my perspective, really nice job that bringing us down to – for people like me to understand. (Ruth), can you take this one?

(Ruth Martin): Sure. So, essentially, the differences that data sources right. So, the chart abstracted measures, you rely on a human intermediary to collect the data and abstract the data into a particular format. And this could be yes or no answers to is there a diagnosis, and that abstracter can actually use a lot of information including information that's documented in paper and documented in the free text form in an EHR, whereas the electronic clinical quality measure, there are two aspects that make it a little bit different, so that you write the focus and the intent are the same.

But the data sources are limited to structure data sources in the electronic medical record that can be accessed to automatically collect that data. And then, the measure calculation itself relies on standard – it's like standard-based representation that hopefully EHR vendors can consume directly, so no one has to implement the logic associated with calculating the measure, multiple times then it's implemented very, very consistently across different EHR products in different sites. I can go into more detail and follow-up questions are welcome.

(Marlin Matoski): It sounds like the explanation was helpful, so.

Woody Eisenberg: Yes, this is Woody. Let me just ask (Ruth), which field consistently provide the viral load data that we need?

(Ruth Martin): Sure. So, we'd be relying on lab test results being an imported into the EHR into a numeric field. And as of I think, the latest iteration of meaningful use stage two, the – this was a requirement that lab test results were reported in a structured fashion. I think that there is – there's definitely some way to go.

So, we're not there yet, where we can say that a 100 percent of providers have this data in place, because it is – it wasn't meaningful sue requirement for stage two. We believe that a lot of progress has been made in that front. And I actually don't have the exact number that we reported and the feasibility's score card but we did look at the data that was reported to the health I.T. committee.

And I stated previously when we were discussing the frequency of medical visits measure, measures where we filled out the score card with the help of a technical expert panel in informatics in using the data that has publicly available but we are actively pursuing testing these assumptions with actual clinical sites and EHR vendors. So, we're hoping that in the next actually couple of months. We will have a lot more data on how these standardize codes and fields maybe use in the field.

Woody Eisenberg: And those data come straight from the laboratory provider, and this updated electronically into the electronic of record?

(Ruth Martin): It could be so that's one of the ways that could get in there. Other way is, you know, that the information maybe faxed in and if there – if it's a PDF into an EHR then it wouldn't be accessible to the electronics measure. But if the data is actually parsed and entered into the electronic medical record, then it would be accessible.

So, the sort of variation – we do expect that there is variation in this process and that's kind of the focus of our additional study and testing of these electronic measures is kind of targeting.

Woody Eisenberg: All right. So, there could be manual steps.

(Ruth Martin): There could be manual steps, you're absolutely right. The other aspect of it does influence the potential performance of electronic measures is the integration, right. So, if that inauguration isn't in place, is so still lab data isn't updated, automatically and/or consumed automatically into the EHR then you would see that manual stuff, how they can happen.

We have had some conversations with folks at HRSA and in some potential test sites as far as their use of lab data and at least some of the sites that use CAREWare are actually, currently importing lab data into CAREWare automatically. We don't quite know exactly to what extent the full population of sites is doing that, but at least some are already in that boat.

Woody Eisenberg: Very good, thank you.

Melissa Marinelarena: And this is Melissa again. If you look upon the screen now, that conversation was around feasibility and that is a summary, if you look in their feasibility and on your own fee, we have a link to the feasibility score card which have a lot of detail. And we try to summarize, I think, with (Ruth) just said as well.

Christy Skipper: OK. So are there any questions or comments about reliability, validity or feasibility, if not, we'll move on to usability and use and the interest of time. We have one more set of measures to run through.

OK. Hearing none, just quickly on – a quick summary on usability.

Laura Evans: So, finally for usability, there's no current uses of the measure that are publically reported or in an accountability program but plan to use an accountability program. Again, sort of reviewed by the measures application partnership for consideration and MIPS that – and in the improvement result are reference of result from the chart obstruction measure, as well.

So, I think the general sense, if you can scroll down a little bit, I think the – preliminary rating on use and usability was high, if I'm not mistaken, I don't have it in front of me, moderate better.

Christy Skipper: OK. And just a note, again there were no pre-meeting and member comments received. Are there any questions from the committee? OK.

Now, we move on to measure 2083 Prescription of HIV Antiretroviral Therapy. Our lead discussants are Jeffrey Hart and Melinda Neuhauser. I don't believe Melinda Neuhauser is on our call. Is Jeffrey Hart on the call?

Jeffrey Hart: Yes, I am, unfortunately. OK. So, I think what I will do is speed through this because these are very much related to the measures that we've been talking about probably skip a few sections because they are rather identical. This is measure 2083, Prescription of HIV Antiviral Therapy – my screen just went blank I guess it's because I'm not touching it.

Prescription for HIV Antiretroviral Therapy and, again, it's from HRSA. The measure looks at the percentages of patients regardless of age, with a diagnosis of HIV prescribed an antiretroviral therapy for the treatment of HIV infection. And during the measurement year, a medical visit is any visit without is the same, outpatient, nurse practitioner, physician or physician assistant who can provide comprehensive HIV care.

The developer rationale is consistent with the other measures based on the current treatment guidelines. So, looking at the numerator, it is the number of patients from the denominator prescribe HIV retroviral therapy during the measurement year, the denominator is of course the number of patients regardless of age, with HIV diagnosis and at least one visit during the

measurement year. There are no exclusions for this measure, it's a process measure.

Any questions? I think this is kind of stepping one step back from the viral suppression to getting the prescription to cause the viral suppression. So, I guess we can move to evidence?

Christy Skipper: Yes.

Jeffrey Hart: OK. The evidence supplied again is related to the clinical guidelines, they did not provide any quality, quantity or consistence of evidence data. And as a result of that, I'll just say that it becomes an insufficient which for a process measure and up reevaluation, it's OK. Are there any questions about the evidence? OK, let's move on then.

So, next is gap and care opportunity for improvement and the disparity section. This actually is the provider performance score for ART treatment, A-R-T. And I noticed that these were exactly the same numbers as the viral suppression rates. So I'm wondering if there was a mistake or if I was reading something wrong in the last presentation.

I don't know if I want to press that but I did think that there was a consistent commonality in the performance for this measure. Did anyone have the other measure they could just quickly look?

Christy Skipper: Yes, the viral load suppression measure, we can switch to that one really quickly.

(Marlin Matoski): So this is (Marlin) on the developer team. If they are the same, it would be an error perhaps a cut and paste error on our part. There is a difference between viral suppression and prescription of antiretroviral therapy in terms of performance.

Jeffrey Hart: Yes, there should be.

Christy Skipper: We'll make sure it's not on our part two. It could have been on our part.

Jeffrey Hart: Oh, it looks like they are different; I don't know what I was looking at.

(Marlin Matoski): OK.

Jeffrey Hart: OK. I think – I think – well …

Christy Skipper: It could because the table is the same, it the same format, so it …

Jeffrey Hart: Right. I was confused about the rate because the rate for viral suppression is actually higher than the rate for antiretroviral treatment. And maybe that's because they use it – might they use the same denominator, don't they? Anyway that was a question I have.

As you can see there has been improvement from 2010 to 2014 about 9 or 10 point increase over that period of time. So we see that ART is being prescribed more often which is good.

The next thing let's talk about is the disparities, they provided in this table a chart of the number of patients in each of the – and in percent of patients in each of the race ethnicity categories. So, we can see that there is a large group of different individuals in the population. They do provide a link to the actual data which is a big – I couldn't get that link to work, I had to go find it.

But to the actual data and it does provide percentage as when we do see that there are definitely disparities. So this is rated as moderate, so we can – I mean unless there is questions.

Christy Skipper: And I'm sorry we're having a technical difficulty with one of the table shown there, right here. It's going to refresh our screen.

Jeffrey Hart: That's the table that provides the number and percent of each category of the race ethnicity group across time.

Christy Skipper: Right. It just disparity stated and they provided I believe it was the same, it was all the same categories?

Jeffrey Hart: Right.

Christy Skipper: Right. We're having trouble with that but it's all the same categories in the – worksheet that you have in your SharePoint, for some reason it's not showing up on …

(Off-Mic)

Christy Skipper: … on here but we can just continue.

Jeffrey Hart: OK.

Christy Skipper: If you have any questions or if there's anything else you would like to add.

Jeffrey Hart: OK. Let's move on then. Next is reliability, this is also a paper-abstracted measure. They look at reliability testing was done using the measures score. And let's see, they use beta-binomial model to estimate the reliability was such as we've seen in the other measure. The median reliability was consistently 0.99 during the – 2010 to 2014 period which conclude that this is a very good – and it has very good reliability. In the end, this is rated high which I would agree with.

Validity, again it's consistent with the evidence in a one which was above the validity testing, was face validity only and as we've had prior discussions it was done because of the fact that they're developing the eMeasures, ECQM as opposed to putting more efforts into these measures. There were no exclusions to the threats on validity and there were no risk adjustments.

The data do provide information that they is variability across providers, the bottom 10 percent of the providers had prescription rates lower of – lower than 29.6 percent. And the 90, top 90 percent had rates of 98.8 – 98.3 or higher. So this is consistent with showing that there is validity to the measure.

On the top of the next page is another table which is not showing up again.

Christy Skipper: We're sorry about that.

Jeffrey Hart: That's the – shows the number of patients with viral suppression across providers. Again, this is about viral suppression not about – oh it does prevent

– present the patients also prescribe ART. So we see that there is consistency actually between an increase in the number of patients who had ART and those who had viral suppression, that probably doesn't have anything to do with the validity but it's good see that.

So because empirical data was not provided, this is rates as insufficient which in this circumstance is fine.

Woody Eisenberg: Jeffrey, excuse, this is Woody. So yes, why isn't that table you just describe to us considered empirical evidence of validity?

Jeffrey Hart: I think it has to do with the fact that it was face only. Otherwise I can't speak to that in perhaps the measure owners could speak to that.

(Marlin Matoski): Could you, Christy, hi can you pull up the table that he was speaking to, or were you not able to open that?

Christy Skipper: We are trying to pull that up, if you could just give us a second.

(Marlin Matoski): So you don't – (Melissa), you're much more poised to talk about this in terms of the different types of validity, that can be tested.

Melissa Marinelarena: Sure. So there, we look at different things for validity. This is the meaningful performances which is the table that they are trying to pull up for is all part of threats to validities. So on the threats to validity, if you have the algorithm in front of you. We look at the different types of threats of validity that are relevant to the measure which include exclusions, the need for risk adjustment, the ability to identify statistically significant and meaningful differences and performance which is what that table is showing a multiple sets of specifications and missing data or non-response. Some of these often though don't pertain to measures, so for example this one, you know, it doesn't need to be risk-adjusted.

So, assessing the threats to validity are different in validity testing. When e look at validity testing, we either looked for empirical validity testing, which face validity is not considered. Again, we rated it insufficient because face validity is the very minimum of testing that we accept. They've just didn't use

the very specific language that we require. It does not mean that the measure is not valid. Again, we look to you as a committee and expert in the field and expert to have use the measure to determine if it does make face validity that the measures score can in fact distinguish good quality from poor quality.

We included all of the work that they did for their definition of face validity which we get lots of varying definition to face validity. So, that's – because it wasn't empirical validity testing, when we look at empirical validity testing it could be either testing of the measure score, which you could look at the correlation between two measures with similar differences or outcomes. And then, we also look at patient level data validity testing which, if we were to have the – re-specified eMeasures that would be really important, because we want to be able to know that we can pull out those data element.

But this – what they did, then we say it's insufficient, we look to you to determine whether it is or isn't. So, that's the difference between that table and the face validity in a very long-winded answer.

Woody Eisenberg: So, maybe I asked the wrong question then. (Marlin), is there a correlation between prescriptions and viral load suppression? Either as a results, from the results of this measure or in a literature?

(Marlin Matoski): Yes. In a literature there is substantially documented evidence that when a person is prescribed antiretroviral therapy and is adherent to that antiretroviral therapy, they will become virally suppressed. We are now at a point in time where, you know, on average if I were to start antiretroviral therapy today, I could be virally suppressed in approximately eight weeks. So there is a very high correlation.

Woody Eisenberg: Right. So, that's if they are adherent to the therapy that's a different kind of a study. For that, you'd need to have either dispensing records, or claims records, or patient diaries. So, the question really is this, is there a correlation demonstrated simply between the prescribing?

(Marlin Matoski): Yes. Do you want to comment on – so I have with me one of our statisticians who helped with the analyses. The correlation is weaker when you're looking

at just specifically the prescription of antiretroviral therapy, why we go with the prescription of antiretroviral therapy is, it is a more readily available data elements?

Woody Eisenberg: Yes. Sure.

(Marlin Matoski): It is rather difficult to get, you know, or pick up data or, you know, patient actual putting of the medication into their body. So that's why we go with the prescription.

Woody Eisenberg: Yeah. Now, I understand that. Thank you.

Jeffrey Hart: I might just comment that I – would number one, we're over our time and I don't know if we want to continue this or we would do this at the next meeting.

Christy Skipper: Yes. So, we just kind of ask if you briefly have any questions or comments about the feasibility and usability for this measure. And would have taken your comments on evidence and in gap, and applied those to the eMeasure. But we won't have time to discuss the eMeasure. So, if there are questions or comments or brief summary that you want to provide on feasibility and usability you can do that now. But we definitely will have time for discussion at the in-person meeting.

Jeffrey Hart: I have no comments especially if we're going to pick that up at the next meeting. I mean, my comment is, these are basically very similar, not exactly the same as the other measures and both the feasibility and usability are very high.

Christy Skipper: OK. Thank you. So, we have completed reviewing all of the measures on this call today. Before we wrap up, I just want to ask the operator to open the line to here if there are any member or public comment.

Operator: At this time, if you'd like to make a public comment, please press start then the number one on your telephone keypad.

And we don't have public comments at this time.

Christy Skipper: All right. So, thank you all for a very successful first workgroup call. Our next workgroup call is tomorrow from 1:00 to 3:00 p.m. Eastern where we will be discussing the sepsis measures and we look forward. If you all are able to attend, you also dial and attend. So, we will conclude the call now?

Male: Melissa, let me just say that one of the sepsis measures hasn't been posted yet, is that correct?

Melissa Marinelarena: That is correct, we have – and that's 0500, we've ask the committee to just review the measures specifications on the project page and direct your questions or comments to the measure specification.

(Marlin Matoski): Correct. So tomorrow's conversation, we'll have the sepsis mortality measure with – and we do apologize for sending it out so late. But we have been – we were still getting – we are working with developers to get some information on this. So, the preliminary analysis for the mortality outcome measure was sent out. So, that will be discussed and the developer will be on the phone and that's a new measure. So, you know, we'll go through the same conversation that we had today.

For the bundle measure, we have not finalized the preliminary analysis yet. So, it will be more having, you know, sort of discussing question, anything like that especially based on experience with the measure. The developers will be on the phone as well, so we can ask them questions directly. And then we can talk more about process. If you are able to attend even though you're not assigned to the measure, we do encourage you too because the entire committee unless you have been recuse from a measure will be voting on this. So, it will be another great conversation and we hope like we had today.

But if not, we will be having all of these conversations again in two weeks. And there will be a quiz at the meeting. So, you will be all experts on this, on the NQF process by the time we're done with this project.

Christy Skipper: All right. So, thank you all for your participation today and we look forward to meeting you at our in person meeting in two weeks. And please reach out to me of there are any questions via phone or e-mail. And so we'll …

Male: Very good.

Christy Skipper: … wrap the call here.

Male: Bye.

Male: Thanks.

Male: Thank you.

Female: Thank you.

Female: I was at the beginning, is it always better to be at the end.

By then, they are just rushing through it. I talked at the beginning and it's always better to do it at the end.

END