NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
March 25, 2014
1:00 p.m. ET

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

Adeela Khan: Hi everyone, this is Adeela from NQF. I'm here with my colleagues Kaitlynn Robinson-Ector, Elisa Munthali and Ashley Morsell, as well as our consultant Robyn Nishimi. This is the Health and Well Being Workgroup 1 preliminary evaluation that we're going to be going over today.

So the way the call is going to work is that we're going to be introducing the measure, we'll read off the measure description and then we'll have the lead discussants kind of go over some of their comments which we've been able to compile and we're projecting on the webinar right now.

I just want to make sure that we have our two developers on the call so Pamela Owens and Patrick Romano should have open lines at this time as well and also Rita Lewis from NCQA. So (Amy), if we could please just make sure that their lines are open.

Operator: Yes, ma'am.

Adeela Khan: We'll refer to the developers of the measure as we move along. So we can go ahead and get started. The first measure we're going to be going over – let's do actually, let's do a quick roll call, so.

Do we have Amir Qaseem?

Michael Baer?

Michael Baer: Here.

Adeela Khan: OK, Ron Bialek?

Eric France?

Eric France: Here.

Adeela Khan: OK. Margaret Luck?

Margaret Luck: Here.

Adeela Khan: OK. Thomas McInerny?

Thomas McInerny: Here.

Adeela Khan: Katie Sellers?

Katie Sellers: I'm here.

Adeela Khan: OK. Amy Minnich?

Amy Minnich: Here.

Adeela Khan: Jacqueline Moline?

Jacqueline Moline: Here.

Adeela Khan: Arjun Venkatesh?

Arjun Venkatesh: Here.

Adeela Khan: OK.

Ron Bialek: Hi, this is Ron Bialek. I was late dialing in.

Adeela Khan: Great. OK. Thank you.

Ron Bialek: OK.

Emilio Carrillo: This is Emilio Carrillo also.

Adeela Khan: OK. Are there any other committee members on the phone?

Catherine Hill: Catherine Hill.

Adeela Khan: OK.

Thomas McInerny: This is Tom. To enter the meeting, we need a webinar number.

Adeela Khan: The webinar information is actually on the agenda.

Thomas McInerny: The 546100?

Adeela Khan: Yes, it's 546100, that's the number you're going to enter in.

Thomas McInerny: Thank you.

Adeela Khan: Yes. So let's go ahead and get started. The first measure we're going to be going over is Measure 280, it's Dehydration Admission Rate. This measure is the number of discharges for – sorry, OK. So the measure is used to assess the number of admissions for dehydration per 100,000 population. It was developed by the Agency for Healthcare Research and Quality and so let's go ahead and have the two discussants go ahead and talk about their thoughts of the measure.

I believe that's Arjun and Michael.

Arjun Venkatesh: I'm happy to go. I was out of the country for the last week and a half so I can only have (inaudible) to go through and review all this and I've submitted a preliminary review this morning. So if Michael wants to lead I'm totally OK with that or I can.

Michael Baer: Well, I don't know. Do you want us to go down through the different questions that were asked in the evaluation?

Adeela Khan: Yes, you can just go ahead and give us your preliminary thoughts about what you thought about the measure.

Michael Baer: Just going down through this, like my colleague I got done this morning. I felt that – I've never done this before so this is brand new to me so bear with me if I may not be as well honed as some other folks who may have done this before. But in terms of the evidence to support the Question 1A, evidence to support the measure focus. In the report that we got there was evidence that – there was sufficient evidence to support the – this incumbent measure and specifically, historically, given that there is a widespread problem with folks who developed dehydration probably mainly in response to gastroenteritis which is a later measure, this measure was put into place and has been reviewed extensively and in terms of Question 1B, in terms of performance gaps, it did appear to be – performance gaps in terms of there was disparities noted by age and also by income.

It's interesting that when I was looking at this since this has gone into place the CDC has developed a national outbreak reporting system which I thought was pretty cool and they had report in there about this particular subject and I put this link to that in my comments that I sent in this morning. I don't know if that would be available to share with the group. But I think that supports the performance gap questions.

In terms of high priority I think this is a very high priority and if you go all the way down to the last question, usability and use, CMS has gone so far as to add this to their value base payment modifier to go into effect in 2015 for groups of provider with greater than a hundred providers in that group. So I think it's a high priority not only – well, I think it's just a high priority nationwide.

In terms of Question 1D, the composite measure, that was not applicable. In terms of the specifications I didn't see any issues with the specifications related to this. There seem to be adequate use of the coding that would capture this information.

Reliability testing I knew very little about. I will just admit that but based off the report that was sent with the information about this measure. There were analyses done which did show that the data was reliable. There was some slight drop in the reliability in areas with less than 23,000 members or patients.

In terms of validity testing, Question 2B, it appears that the construct validity testing did show a positive value reflecting the measure was valid and threats to that validity, there was some information about discrimination and calibration and there was moderate discrimination for both – there was a moderate result for both discrimination scores and calibration scores.

Again, on Question 2D, this was not applicable. This apparently is not a composite performance measure. In terms of feasibility, this is highly feasible as this information is gathered through electronic sources, HCUP is one of those sources and also those who are coding these are – other than the person obtaining the information so they felt this was highly feasible and in terms of usability and use there – you know as I mentioned it's interesting that this is being used by CMS in their new value based payment modifier to go into effect in 2015 for groups of greater than 100.

There are other multiple public uses for different states to allow the public to view information about this condition so that's pretty much my overview of that.

Adeela Khan: Thanks. That was really well done. We can actually – we'll add in – we're going to be compiling all of the notes that you all are talking about and we're going to be sending that out to the committee prior to the in-person meeting so we'll have that link included as well.

If you included it in your evaluation then we'll include that when we compile all of the three workgroups together and hand it out to you.

Michael Baer: OK, I did include that.

Adeela Khan: OK, great.

Arjun Venkatesh: Adeela, should I have (inaudible)?

Adeela Khan: Sure, if you – yes.

Arjun Venkatesh: So I mean I largely agree with what Michael just said. I think that there are – what I'm going to try to focus on maybe is the places where we may either disagree a little or where I think we probably need a little …

Adeela Khan: Arjun?

Arjun Venkatesh: Yes.

Adeela Khan: Thank you. That's actually better. We couldn't hear you so if you could speak just a little bit louder, that'll be great.

Arjun Venkatesh: OK. This is OK right now?

Adeela Khan: Yes, thank you.

Arjun Venkatesh: I was largely saying that I agree with Michael I think across the vast majority of these domains, these measures in terms of all the PQIs have been around for a little bit. This is kind of now the second iteration of them that includes more of risk adjustment, indirect standardization and I think that will show us the evolution of the measure.

I think the place where we probably need the most discussion in terms of the group is around 1A and 1B. I would agree that just kind of based on what Michael said that in terms of whether or not this is a high priority area, I think that dehydration is considered that based on the references they have.

It's related to a lot of the high impact conditions that have been identified by HHS and it can be fairly high resource use. I think the usability of this measure and its feasibility are both very high in terms of the fact that it can be pulled both from administrative claims make it peaceful and then the usability obviously is by meeting QIO state collaborative part of payment programs.

What I wanted to focus on thinking a little bit about is 1A and 1B. 1B relates a little to the validity and so if Patrick and Pamela have other things to add I think it'd be helpful.

When we think about the evidence to support the measure focus, there are variety of references that are in the (myth) form that are submitted that talk about how dehydration is epidemiologically a large issue and how hospitalizations carry high mortality. What there isn't great evidence around or what there hasn't been really cite much around is preventability of dehydration hospitalization. And as I look through what they kind of had cited within the actual submission form, when we look at the guidelines to support the measure focus, the interventions around preventing dehydration hospitalization are all either level 4 or level 5, meaning non-experimental design or kind of expert consensus.

And so one of the questions I kind of pose to the group is to what degree do people feel that hospitalizations for dehydration really represent a community level in this case for this measure the data will (inaudible) at county level. So a county level phenomenon that carries preventability and sort of related to that is one of the recommendation and the rationale for measurements that supported this surveillance for dehydration.

And while there's definitely some studies that suggest that you can use tools to find higher risk individuals, the elderly, those with cognitive impairment or just dementia, there's not really much of an evidence focus to suggest that – sorry about that, there's an ambulance outside me here. There really isn't a strong evidence base to say that implementation of a surveillance program or a county level surveillance can lead to reduction in this rate and so I think that while there's definitely, I would say strong evidence of the measure area, I'm not so sure about the idea of preventability of dehydration hospitalization.

Around the performance gap I think that there's certainly large county level variations that they have shown and so that would suggest that there is performance variation. My concern though is that the rate of these preventable hospitalizations dropped from 155 per 100,000 population in 2007 to 216 per 100,000 in 2011. So in four years there was an almost 40 percent

decline in these hospitalizations and to me I don't – can't think that the evolution of care, especially in the absence of these measures either being used in any payment programs, master public reporting efforts in that time to reflect that.

So what my fear is, is that either there's some coding shift which is something that we probably saw with some of the anginal measures around PQIs. Or that this is a reflection of changing in coding patterns regarding care sets. So we're these previously in patient hospitalizations in 2007 and now these patients are being managed in kind of prolonged (inaudible) visits or observation visits that may be absent from the measures or this true quality improvement.

And the reason I think this is a big deal is that if we're showing a 40 percent drop in this measure over four years and a large percentage of that is because of more measure validity based type reasons in terms of what's being captured in administrative claims then this may not be a very good measure to use to compare counties because this is largely just the variations in observation care coding patterns.

And so – and I know the developers alluded to that in their submission form that some of the outpatient visits were nicely captured and so it might be worthy of some discussion kind of around that. But I think those were kind of my bigger picture concerns about this measure kind of going forward.

Adeela Khan: OK. Thanks a lot, Arjun. Is there anyone else from the committee who wanted to provide some insight into the measure?

Jacqueline Moline: This is Jacque Moline. You know I second your – I thought that that was a huge drop as well and I was thinking it was exactly what you were saying. I think it was raised that that might be one the challenges which is someone's in a 23-hour stay in an observation unit, it's not being coded as inpatient. For all intents and purposes, it is like an inpatient but it's not coded that way now and that's where folks are getting the treatment so the measure validity won't be – you won't be – if the goal is to prevent hospitalizations or to prevent hospital visit for dehydration, are we adequately capturing it?

If the goal is to say that someone need more than 24 hours or 48 hours or whatever it might be, then you are going to see it. It's reflecting more change in practice patterns within the ED.

The other question that I had when I was looking at it was I was thinking of different sources of dehydration and maybe it's because I was also looking at the pediatric one which is virtually always related to GI. That I wondered, is it really preventing dehydration or is it preventing dehydration due to gastrointestinal illness because that's really what they're looking at because there's other positive dehydration comorbidities and so on in chemotherapy and/or that would require in some cases hospitalization. I wasn't sure how that was being captured.

Arjun Venkatesh: Just two points I think based on what you said I didn't mention when I was discussing my review is one is that the way the adult measure is constructed is you're absolutely right. Some proportion of this is kind of diarrheal illnesses, the other proportion of this is probably folks with chronic condition that either end up with dehydration as a result of changes in PO intake and interactions with medications or result of kidney disease and so, there's not really data around – I don't think that I saw in terms of how each of these populations makes up the population of the measure but all that diarrheal is part of this as well.

And then the other thing I was thinking about as you mentioned about observation cases is the goal of the measure is to reduce community level acute care visits for dehydration, it's the ones that end up with ED visits only or observation visits that might actually be the ones that are most preventable versus those that require more prolonged hospitalizations are likely folks with more comorbidities and a variety of other factors that may make their actual acute care visits less preventative. So in some ways we may not be capturing what we want to capture.

Jacqueline Moline: Right. So it's almost like two different things that would be captured and I would surmise that the 40 percent drop is purely related to the fact that it's treat and release versus we didn't decrease by 40 percent, especially in an aging population.

Ron Bialek: Hi, this is Ron Bialek. I had a couple of comments. One was that the measure seemed to be constructed to apply to 18 and over population and the evidence suggests that the real problem is 65 and over. I'm just sort of wondering if the measure is going to be sensitive enough to make that for it to continue to capture what we're trying to capture here.

Second comment has to do with performance seemed to be far better than the last and there was really no discussion of that and I'm wondering if the measure is going to get to that at all or if there really is an explanation for that.

The third comment is about the database. My understanding is that the present on admission data element is added in the 2006-2007 time period and that the data for – that were presented were 2007 to 2011 and when there are new data elements added off and there are some definitional issues, reporting issues, et cetera, and so it's hard to know if the changes that occur over that period of time are related to data issues as mentioned before or something else.

But like you said I don't really know the dataset particularly well. I was just looking at some of the literature around that and it seemed to be the present on admission was newer.

Adeela Khan: Thanks Ron. Is Patrick or Pam on so they can answer some of these questions that were raised?

Pamela Owens: Both Patrick and I are on and this is Pam Owens from AHRQ. I am the senior lead on the AHRQ quality indicators and Patrick will be able to fill in most of it. I want to speak to the last point first which is this indicator does not use present time admission as a data element. You are correct in saying that present on admission is a data element that's come up within billing data within that time period. However, in 2007 is when Medicare required POA reporting and it wasn't until 2008 that there was a payment penalty in place around whether or not a hospital used POA.

So all that to be said is yes, there was the state element however it doesn't relate to this particular indicator. It is used in some of the AHRQ quality indicators but not this one.

The second point, your question about on seeing it in the outpatient care. Certainly, for your review, your in-person review, I can run some numbers on the outpatient as Arjun knows. The other hat I wear is I am the coordinator of the outpatient data for these data projects so I can tell you what the utilization is in the ED.

Preliminary analyses unrelated to this, I would suggest that in fact we are seeing that. Patrick can speak to the clinical nature from that.

Patrick, do you want to chime in now and I can speak more on the data side if need be.

Patrick Romano: Sure, I'll just address a couple of points. Very good discussion. With respect to the treat and release phenomenon, yes, I think most of us from the clinical side would say that the major reason for the drop that we're seeing is a change in practice pattern. I wouldn't say that that's necessarily undesirable or evidence gaining. But it is a change in practice pattern.

The hospitals are generally not good places for people to stay if they don't need to be there. So one might argue that treating these patients for 12, 24 hours in the ED or in a facility connected to the ED may be better care than admitting them to the hospital but that's arguable of course. But I think we would all agree on that.

The other thing just to clarify is that it is the intended indicator to capture dehydration that is related to acute conditions and not chronic conditions. Now the coding rules stipulate that if the cause of the dehydration is known then that cause should be coded as a principle diagnosis. And the dehydration should appear in the secondary field.

So by virtue of the architecture or the indicator, we're inherently capturing patients who were presented with gastrointestinal illness with dehydration as a secondary diagnosis or patients with idiopathic, if you will, acute dehydration.

So if the dehydration occurs for example in the setting of chronic renal disease and a patient on dialysis who may have been over dialyzed, that patient will have a renal diagnosis, just the principal diagnosis would not be captured by this PQI.

Similarly, a patient with heart failure who was over diuresed would not be captured by this PQI. So there is a focus on these acute conditions based on the premise and I'll admit that this is an arguable premise. But based on the premise that these – the kinds of conditions where the patient may be more safely managed out of the hospital perhaps with oral dehydration or just a few hours of intravenous rehydration in the ED.

So I think those are the clinical issues that I wanted to address, but certainly open to other questions.

Adeela Khan: Yes, we have a couple more minutes, so we can continue talking about this measure. Does anyone else have anything else that they wanted to raise?

Amir Qaseem: So Adeela, this is Amir Qaseem calling from the American College of Physicians. I think Arjun it was, I think he hit the nail on the head about regarding all the key point that we need to really keep those in mind. And I just couldn't – actually, my own issue is I just couldn't get pass in terms of what the evidence was presented. Just what is good clinical practice does not necessarily mean that we can convert it into a performance measure and measure a physician's performance against something that is just good clinical practice.

So that if you look at some of this evidence, actually I was quite surprised that this was an AHRQ sponsored performance measure, because if you just – AHRQ uses its own criteria, the evidence that was presented, they'll not make it in terms of their own guidelines. And I think that this measure falls apart what was presented in terms of the basis of evidence. Many of them is taken from the National Guidelines Clearinghouse but they are not really clinical guidelines per se according to all the definitions that we have been talking about.

Actually, AHRQ has been talking about and I think until we have some of that, it is good clinical practice. I'm not going to question that part but that does not necessarily mean that physicians should be measured and that information should be made available to public that what has been happening or not happening in terms of taking care of patients with dehydration.

So that I think that has been my biggest issue. I just didn't get pass basis for the recommendation, basis for the performance measure.

Patrick Romano: Well, this is Patrick. We need to be very clear that this is not a physician performance measure or a hospital performance measure or a physician group performance measure.

This is a performance measure that's only intended for applications to a population and it's really intended as a measure of population health and the effectiveness of the overall health system for that population. And it says here that physicians provide is only one component of the broader public health infrastructure that affects the rate of this indicator.

So it's really intended for use by public health agencies, by local agencies looking at comparisons across population, typically geographic population, just to be clear about that.

Male: So Patrick, a question for you that I have is that would you say that the evidence that was presented meets the threshold that's set by AHRQ when it comes to clinical evidence, the quality of clinical evidence. Where would you place it? Is it high, moderate, low?

Patrick?

Patrick Romano: I'm not sure I understand the other criteria that you're applying.

Amir Qaseem: So, or any evidence, I mean you have some evidence that's presented under clinical practice guidelines recommendation, some basis there, some things that are presented under, you know, what goes into the evidence part.

How would you classify that evidence using AHRQ's own threshold when it comes to assessing quality threshold. You can use whatever format you want. You can say class 1, class 2, class 3. You can use high, moderate, low, how would you want to present it? Where would you put that evidence?

Pamela Owens: So, to be fair to Patrick I am not sure he is aware of – I actually downloaded on the clinical practice guidelines and it was not necessarily supposed to be taken into context but I think you're reading it as much as there is some evidence out there regarding guidelines and I believe I don't have it open in front of me. I think it largely related to nursing homes although I may be mixing up my PQI.

So Patrick may have to get back in touch with you. It would be my fault if I have misrepresented the indicator.

Amir Qaseem: And that's fine. My whole point is I think no one on the call is going to disagree with the fact that we need to have evidence before we can have performance measure and I think that's the point I was making in here.

Catherine Hill: This is Catherine and from a nursing perspective, you know, I'm new and so I'm learning from you all. These is very robust discussion but from a nursing perspective, we think about the determinants of health and hydration and certainly one of those things that we feel like is dependent on individual behavior and that it is very amenable to intervention, education, community outreach. And so I'm wondering if that influences the opinion today.

Arjun Venkatesh: This is Arjun. I would agree with what you just said. I think probably where this would help us in terms of between now and the next meeting that's all about this measure, is that there is certainly some phase validity to that statement, that public health and a variety of community based interventions around surveillance intervention is going to improve hydration status.

And I completely agree with kind of what Patrick was trying to guide us towards which is saying that, you know, both of our committee (inaudible) these discussions to really think about this as a community level measure where we think about how a variety of public type interventions can improve

the outcomes, where I think I have a little bit of challenge there is in terms of understanding the degree to which variation in the performance score.

So for example I live in New Haven County and Fairfield County is right adjacent to me. And to know that when we have higher number of dehydration related admissions in our county than in their county, that represents a truly preventable number first of admission and that if preventable, they are amenable to community level intervention.

And I think that those are the two steps where we probably need a little bit better. I personally don't know that literature very well but needs some more literature evidence to believe that those linkages are true in terms of this outcome because otherwise, we may simply just be measuring a variety of other also related social determinants of health around socioeconomic status and a variety of other things between counties that are not truly kind of preventable through community level intervention.

Margaret Luck: This is Margaret Luck. I agree with you and I guess my fundamental question is, if we see a decline in this rate, what does that mean? Does that mean dehydration is being prevented at the community level or does that mean people are being treated in the ED rather than admitted?

Michael Baer: Well, we do have some data that we can use to empirically explore that question to inform the committee discussion at the in-person meeting.

(Off-mike)

Adeela Khan: Yes, I think that will be really helpful for everyone if you could just kind of pull together like what we've talked about today in the discussion and bring that to the in-person meeting just so it will be helpful for this group and the larger group also so that we don't, you know, lose sight of the conversation.

I was going to say that in terms of evidence, this is an outcome measure and so I just wanted to put up on the screen really quickly again the evidence that we are looking for in terms of for health outcomes, that we're looking for a rationale that supports the relationship of the health outcome to at least one healthcare structure process intervention or service.

So this is from the measure evaluation guidance that's also posted on SharePoint for all the committee members to take a look at. But again, I just wanted to put it up there just so the refresher for everyone so that it's clear what kind of evidence we're looking for.

So I'm going to actually move on to the next measure. We do have four more measures to go over during this call. The next one is going to be urinary tract infection admissions rate, that's PQI12 and that measure I believe is the lead discussant is Ron and Eric.

So one of you …

Eric France: Hi, this is Eric France, first time caller, love your show. This is a very cool conversation and Arjun, as you were talking about staying high for those first 1As, that was certainly sort of how I was thinking around this measure.

Ron, are you OK if I just jump in?

Ron Bialek: Go ahead.

Eric France: Yes. So again, forgive me as I learn this whole process but this is a measure that's part of the ambulatory care sensitive condition group of questions.

And so for my perspective, I'm framing it from a county that's saying do we have the right ambulatory care resources in place so that people are treated early with the urinary tract infection and do not end up in the hospital.

And in general, the rates of the UTI hospitalization per hundred thousand population has actually been going up from 2007 to 2011 from 160 to 180 per 100,000 population over the age of 18. And like the last measure, it particularly seems to be at increased risk for persons over the age of 65 and annual risk of 200,000 in general but eight in a thousand for people over 65.

The increasing rate might be explained by either increased ED use for primary care which might then lead to hospitalization. Also, other thoughts are that it's due to the increasing prevalence of diabetes and obesity and aging population

and potentially resistance to oral medications requiring higher hospitalization rates.

The rate itself is developed from HCUP as well and from the ideas of reliability and validity, the review seemed reasonable, the data's element seem to be fine. I went back to that broader question of understanding the evidence that this measure would be impacted with improved access to primary care knowing that the rates of the UTI hospitalization have gone up over this five-year period.

If Obamacare improves primary care access will we see that this is a sensitive measure and the rates actually go down, I'm not sure that the evidence that was presented didn't really show that there's a nice evidence base showing that there is that impact to be expected. So I think because of that sense of absence of information around that I was a little hesitant about saying this was important.

I think too, I wondered whether measures are ever applied to specific age groups rather than the whole population. Would this be better if it were a metric for people 65 and older rather then the whole population? And there was a statement too that only 10 percent of UTI hospitalizations might actually be preventable through primary care access, so that made me then wondered about the preventability as well.

So I think I'll stop there and see what Rob wants to add.

Ron Bialek: Hi this is – it's Ron, but …

Eric France: Ron, excuse me.

Ron Bialek: That's OK. I concur with everything that you had said. I sort of wondered about again the age grouping issue. There was a little bit presented on preventive measures but really nothing presented that I can see evidence on how to impact the use of the preventive measures.

Then with regard to the primary care issue, there were a variety of activities at the community level that can impact this that access may not at all be the issue and the evidence really wasn't presented there but looking at other areas.

For instance, the ER becomes sort of the customary place somebody goes, the proximity. So that maybe insurance access that way but it may not – people still may be using the ER. And then secondly, there's a lot that impacts the emissions that are related to the diagnosis treatment that occur outside of the hospital setting as well.

Last which is less comment, which really is more of a question, I still don't fully understand the HCUP data in terms of the capturing of the diagnosis at admission versus the diagnosis of discharge. And it looks like the POA was not part of this, so that's not really issue but I'm still trying to understand how the – that those data are captured at the admission point.

So maybe the developers can just clarify that for a moment.

Pamela Owens: Sure, this is Pam Owens from AHRQ. On the HCUP data it is largely a billing database, there's a couple of states that provide us discharge abstracts. But a billing database is similar to a claims database in that they use the Uniform Bill 04, which has very standardized guidelines from the National Uniform Billing Committee about how things are coded in the order that they're coded and why they're coded.

And since it is a discharge bill and the reason I bring up the claims database, so, you know, if the Medicare uses this indicator, for instance, they're using a claim, that means that it's adjudicated, they've taken the bill, they've figured out what parts they're going to pay for and then it becomes an adjudicated claim. All that to be said is the data elements themselves, they're standard whether you're looking at the HCUP or whether you're looking at Medicare, it would be database that is collected for reimbursement purposes.

And then it's used to do some reporting on terms of utilization and, you know, extrapolate the quality of care. So what does a diagnosis mean? Well the principle diagnosis or what you might understand to be a first list of diagnosis, but in an inpatient – on an inpatient bill that principle diagnosis has a distinct

meaning, which means after all is said and done and considered, what is the reason for this hospitalization?

And that is the only thing that can go in that, what would be the DX1, the first diagnosis. And we call it, it has a specific name, it's called principle diagnosis. All of the other secondary diagnoses they don't have a set order of which they appear but they are to be any of the diagnoses that required some resources on the hospital, this is the hospital bill, it's not a physician bill.

And so we need diagnosis that then increased resource utilization on the hospital side. It could be a complicating condition, it could be a comorbidity. Those are the kinds of things you would see on a secondary diagnosis.

Does that help?

Ron Bialek: Thank you. Yes.

Adeela Khan: Do any other committee members have anything to add?

Jacqueline Moline: This is Jacque Moline. You know, one thing I was a little troubled by – troubled is probably too strong a word. But it seemed like the measure of hospitalization for urinary tract infection is different than pyelonephritis which understood is part of the urinary tract. But, that one doesn't seem – to me it's already gotten to a level that most of you are actually know as I'm talking, it means that maybe some – it should have been caught earlier so it didn't progress there. So maybe that's what we should be looking for.

I was just thinking that they were almost like two separate issues but maybe they're not as I'm talking and I just convince myself that I should shut up.

Adeela Khan: No, (inaudible).

Jacqueline Moline: I mean, I don't know if that struck anyone else. It's just that they seem like such different things to measure. And I don't – grouping the two of them together just didn't seem to make a lot of sense when I was looking at their face value but maybe it does because the pyelonephritis should be the preventable, which would lead to a hospitalization and a simple UTI shouldn't.

Male: I think that's right.

Patrick Romano: Yes. This is Patrick Romano. Just to clarify your idea, the indicator numerator does include both pyelonephritis and lower urinary tract infection. And part of the rationale for that is that there's an underlying presumption that if somebody gets admitted it for a lower urinary tract infection, it's because they probably had features of an upper tract infection or because the patient has some underlying, you know, comorbid illnesses that make it, you know, more appropriate to admit them to a hospital.

But effectively I think it's difficult to draw that line between upper and lower tract infection.

Jacqueline Moline: You know, I realize that and I did look at the exclusions and the inclusion when I was looking and I think actually when we're thinking about it's, you know, ambulatory sensitive condition. It really would be, you catch it early enough, you prevent the infection – you prevent the hospitalization. So you have to include those. Sorry to convince myself.

Eric France: There's also – this is Eric. There's – as you sort of sit in the question for a bit, you think about how different urinary tract hospitalization is for an 80-year-old versus a 40-year-old as a condition and it's preventability and it just makes you wonder about more age-specific metrics and banning the metric because the community interventions might be quite different from – for people who are now seniors and coming in with urosepsis. And someone whose 40 whose got sort of the UTI that didn't get treated went on to pyelo and then they're in the hospital.

Just something to sort of file away maybe for further conversations about where we want to consider conditions might be summarize to greatly when they're such a broad age span.

Thomas McInerny: This is Tom. (Crosstalk) particularly because someone over 80 is more likely to develop all kinds of problems while they're in the hospital, secondary complications, et cetera, than someone who's only 40. So they would have a higher morbidity and mortality if they were admitted versus a younger – much younger person.

Adeela Khan: OK. If there is no further discussion then …

(Crosstalk)

Patrick Romano: Could I add a comment?

Adeela Khan: Sure.

Patrick Romano: Yes, this is Patrick again. I think there's no question that that's true from the clinical perspective and we have in fact considered and discussed in the past with our AHRQ's clinical advisory panel about whether to truncate the upper age for this indicator and several of the other PQIs at, for example, 75 for the reasons that have been discussed. And there's no right answer but, you know, we're certainly open to the input and discussion from the steering committee.

In the past, NQF and others have always considered us or encouraged us to apply the indicators as broadly as possible with as broad a denominator as possible to cover as wide a population. But, there – I think clinically would make sense if there's less preventability as you get up to the higher ages above 75 and 80. It's just hard to know where to draw that line.

Adeela Khan: Thanks Patrick. So let's move on to Measure 727, Gastroenteritis Admission Rate. The lead discussants for these are Tom and Amy.

Amy Minnich: Tom, do you want me to start then you can supplement?

Thomas Mclnerny: Go ahead, Amy.

Amy Minnich: OK. And I can echo on my colleague in that this is a new rodeo for me so I appreciate your patience. This measure is again a preventive one gearing more towards pediatric population, does have some relationships to the dehydration admission as well.

As far as evidence to support the measure, it seemed like that was clear. There was a number of clinically based trials that did support the effort of what this was trying to do and to reduce complications resulting from DI admissions.

It is ambulatory sensitive condition which seems to be a theme across all the measures that we're looking at and that can it be better managed in the outpatient setting. As far as the performance gap, it seemed like there was some disparity that was identified in lower economic ZIP codes. So certainly, that would contribute to the measure itself and that there is variation across providers that could impact the quality of care. So that seemed like that criterion was that in my opinion.

Is a high priority issue, giving that one in every 50 children has some type of an acute admission during their childhood related to GI situation. So it does have the potential for high resource use in a hospital setting versus the outpatient. So again, as far as where I would rate that I would say it would be deemed in a high category.

Specification wise, it seemed the measure had been tested and confirmed reliable and valid. Since the last update there were changes made to look at the reference data from 2010 population stat, as well as 2012 population estimates from the census data and also fiscal year coding updates. So that seemed to meet the requirement in that area. Both reliability and validity did take into account risk adjustment data. And so, there was a strongly reliable and valid indicator in both of those areas.

As far as threats, again, it seemed that the exclusions or with anything applicable there on risk adjustment, it did consider gender, age, poverty by county of residence. As far as statistical significant, there was noted better benchmark data in larger population sites than smaller population sizes so that it could potentially have an impact to the validity of the data.

Multiple data sources was not applicable nor was missing data. As far as feasibility, the data that was used was from electronic claims submission and the coding occurred by another person other than who was actually analyzing the data. So, feasibility did not seem to be an issue in this area.

For usability and use, currently ARHQ is using the data, as well as three states that were named throughout the content of the material including Connecticut, New York and California. In the result of future plans, as far as QI and

benchmarking and going forward was a significant decline since the measurement began to be looked at since 2007. 121.5 per 100,000 down to 67.5 per 100,000 in 2011. So overall, those are my thoughts.

Tom, I don't know if you want to add.

Thomas Mclnerny: Perfect analysis. I don't have much to add. You know, I think this is interesting and (inaudible) sensitive (payer) measure because this measures a couple of things in particular. One is the use of rotavirus vaccine and, you know, how much percentage of the eligible population are getting vaccines with rotavirus vaccine and two, how effective the primary care physicians are in helping children to get rehydrated using oral rehydration solution. And so, those two factors are pretty critical in reducing the admission rate.

There are a couple of other factors that might be a bit confounding and that is one of the key, so we end up going to the emergency room and being rehydrated and sent home successfully. I think that's still good because at least they're not being admitted.

And then the other confounding factor that we've talked about already and a couple of the other measures is the business of the short stay units that's becoming more common and that in many cases the insurers including I think Medicare are saying that they have to be in the hospital for at least 24 hours and then that has to include past midnight and that kind of business.

And in many cases, (inaudible) may get admitted even to the hospital but they're admitted for 16 or 18 hours and then they're sent home. So that's not counted as an admission in some instances but are maybe counted on admission or another. That can be a little bit confounding along with the storage units.

But overall, I do agree this is a very important measure to measure and all the comments you made about reliability and validity and feasibility or right on as far as I'm concerned.

By the way, I just wondered, I think Amy and I both did respond to the surveys on this, however, when you hit finish it sort of just disappears from

the screen and I guess we have to trust somehow whether we hit finish, ends up on your computers as the National Quality Forum. There's no sort of confirmation that survey was received.

Adeela Khan: Yes, we received it actually. We've heard that from several people. We're actually exploring some IT options to kind of give you a sort of receipt letting you know that we've received your survey. But we did receive it and we're actually projecting the feedback that we got as of yesterday 3 p.m. on the webinar. So it's there.

Thomas McInerny: Good, thank you.

(Crosstalk)

Amy Minnich: Adeela, I can't see that on the screen and I don't know if I'm the only one, maybe it's a blocker on our system site here.

Adeela Khan: You can try refreshing your screen. If that doesn't work, then we'll be, again, we'll be compiling all of the feedback that we're getting today and we'll be sending it out to you within a week.

Amy Minnich: No problem, thank you.

Jason Spangler: This is Jason. If you go back to that committee preliminary measure evaluation, the survey screen and go to show all responses, it should post your response to confirm that it's been completed. So that may help people to submit.

Adeela Khan: And you know it's there.

Jason Spangler: Yes.

Adeela Khan: OK, thank you. Was there anyone else from the committee who wanted to say anything about the measure?

Pam or Patrick, did you have any response?

Pamela Owens: I don't have any particular response at the moment other than just looking at the comments that were scrolled up. I have to apologize. If I included language around POA to mislead reviewers, I highly apologize for that. I submitted over, I think, 15 indicators for maintenance endorsements in the last two months and they're applicable to some of the indicators and not to others so I apologize if I confused you with that language.

Ron Bialek: This is Ron Bialek and I'm probably the one who is responsible for that. I think I was more confused. Thank you. And we will follow up on the ED treatment question which I think is a very, very good question.

Adeela Khan: Thanks, Patrick.

Pam, this is Elisa from NQF and one of the things that I just wanted to reiterate something that Adeela mentioned earlier, we will give you and Patrick an opportunity to make any clarifying correction to the POA might be one of them prior to the in-person meeting. So when we send out the summaries of the workgroup calls, we make sure that that correction is noted in the summaries.

Adeela Khan: Thank you so much.

Eric France: This is Eric. I think we're not getting a little hang is just me thinking about these measures as outcomes that reflect something about the care infrastructure of a community and how does a new vaccination program may account for a lot of the reduction here of these admissions play into that, she's just sort of resetting the whole bar and restart again now with the baseline of 65 per hundred thousand and then think about ambulatory settings where we can manage these folks.

Is it a validity problem when the reason for the drop is a new scientific product rather than a healthcare system change?

Adeela Khan: I believe that's for Pam and Patrick.

Male: I'm sorry. I don't think I quite followed the question.

Male: These population county-wide population metrics is basically lower, that's good. If the rates go higher, that's bad in a simple way. And the rates are to go lower if we provide more accessible care or manage in such a way so that people don't in the hospital.

And now you have a new technology which is a vaccine that gets put into place in a population. And the benchmark gets sort of reset from, I can't remember the number, 120 down to 65. And how much of that is really the new technology coming in rather than a change in access to care.

And so …

(Off-mike)

Male: Is that important as we're thinking about which of this we endorse and how we endorse them.

Male: Well, I would say that if you look at the trend over time or the population as a whole, for example at the national level, that certainly you'd have to attribute a large portion of that drop to the introduction of the rotavirus vaccine.

But I think what we all know from our clinical experience is in fact nation programs have variable penetration across the population and some subpopulation are vaccinated higher rates than others, some are vaccinated earlier at higher rates.

And so, as we look at any point in time, as we look at the variation across communities, I think some of that variation is likely to be due to variation and to penetration of rotavirus vaccination.

Male: Yes.

Male: So this is really – so I think there are two different interpretations perhaps from looking at the overall trend nationally over time versus looking at a particular year across geographic or communities. Does that make sense?

Male: I think so. And then of course you can argue that the new technology is primarily administered in ambulatory care setting so they are ambulatory care

interventions to vaccinate probably 90 percent of the people who receive this in the first six months of life.

Male: Exactly.

Male: Thanks.

Female: And HCUP staff have done with NIH staff – have done quite a bit to look at the impact of the rotavirus vaccine on these hospitalizations. I'm happy to supply that article or articles to you.

I think the other thing to be aware of in terms of at some point, I think there is enough community variation in as you all were just pointing out in terms of the delivery of this vaccine and people accepting the vaccine that there are still reasons to think that this is a good indicator.

But we do on an annual basis review what's going on with these indicators and if at some point you feel like, you know, we've hit the threshold, it may be time to retire an indicator. I'm not suggesting not at all for this indicator. I'm just saying that your point about, at what point do you stop, is an annual review that we do internally and of course that's part of what NQF is doing by having measures come back every 18 months of the three years.

Male: Thank you.

Adeela Khan: OK. Let's move on to 728, Asthma Admission Rate. The leads for this one are Margaret and Jacque.

Margaret Luck: This is Margaret hoping that Jacque will walk us through it.

Jacqueline Moline: Here I was pausing hoping you would because I (inaudible) because it's – I'm just glad I didn't get any (teeth) measures but I'm, you know, an internist so I was thinking of goodness gracious. But I'm happy to start and please feel free to chime in and like many on this group, this is a new kind of review for me. Give me regular grants and I'm happy, but this is fun.

So this is looking at admissions in pediatric population and excluding folks with inherent respiratory problems like cystic fibrosis. It seemed like there

was a good measure and this is a – basically, it's (inaudible). It was originally endorsed. It was most recently endorsed and actually originally endorsed three years ago. So this is basically taking updated census data and looking at it from that standpoint.

I think the performance gap was well described. The rates have actually been fairly stagnant since 2007 when the rates were 115 per 100,000 and now in 2011, they're 119. So clearly there is room to move and it is age-sensitive with the lowest – the youngest children being most effective.

There is a gap between some levels and also some reasons people in the West are much healthier than those of us in the Northeast which seemed to be in every single measure. It makes me a little scared to be practicing in the Northeast.

But there was really, I think in terms of a high priority, the admission rates for low income and minority children is really, there is a tremendous disparity and it should be a measure.

And also, again, with the national variation and it may be related to obviously environment plays a key role. It's avoidable with proper care or early recognizing of conditions which should be managed from an ambulatory setting.

It was well defined. It was updated with the recent census which was leading to the specificity. The reliability testing was excellent and you know, I think it's really a critical measure because this is one that was very well described, have very – have strong reliability and validity testing that had been provided to us with correlations where they would be expected or coefficients where they would be expected.

And that's basically it. It's one of those, it should be managed from an ambulatory standpoint and the hospitalization rates are there. It should be measured as population based outcome to see how we could be improving the health and the population at large.

Margaret Luck: This is Margaret and I agree with what Jacque has said very much.

The only concern I had about that link between the outcome measure and preventable ambulatory level of care interventions relate to the underlying burden of disease could be different in different areas or it could be changing over time. There could be secular trends in childhood asthma.

And I'd like to see that kind of acknowledged in the section on the evidence to support the measure which is if I understand it correctly, Adeela, really the only component that's required for an outcome measure, is that right?

Adeela Khan: That's correct, yes.

Margaret Luck: OK. I did notice that a lot of additional non-required information was included and then it seems kind of spotty like in the document evidence sub criterion 1A, they noted that this is a clinical practice guideline and then it says, go ahead and complete the two sections about the recommendations and the evidence from the review of those recommendations.

And they went ahead and made, had links to the clinical guideline recommendations and then they failed to provide any summary of the body of evidence. So it was kind of – none of it was required but what they did put was not complete.

Adeela Khan: Does anyone else have anything to add?

Thomas McInerny: This is Tom. You know, I agree this is a very important measure. However, in the area of validity, I think there are some confounding factors that are somewhat on the health systems as a whole. They're controlled. I don't know exactly what you would – how you would characterize compliance or failure of compliance by parents particularly to administer inhaled corticosteroids and other preventative type management.

And then we get into the problem of kids with asthma being exposed to secondhand smoke in apartments even if we've been able to convince, you know, some primary care physicians had been able to convince to parents about the smoke. So if there's smoking in their apartment by other people, the

evidence shows clearly that that smoke does permeate all the apartments in the building.

And then you get into the business of cockroaches and other kinds of problems like that that can aggravate asthma and, you know, those, I don't know, I think they're kind of beyond the healthcare system through control. And so, I think some of the confounding factors that we have to think about when we look at the asthma admission rate.

And then again, this is something, the admission rate can be influenced by either certainly in the emergency room or short stays, you know, in our hospital. And we do have a number of children who get admitted with asthma and progress rapidly to be able to be discharged within 16 to 18 hours so they don't make a "24-hour hospitalization" but they actually were hospitalized.

That's in the form of problems to deal with.

Ron Bialek: This is Ron. You know, I'd like to follow up on that comment. The points are well articulated. Yes, this is a condition where the primary care access and the primary care "treatment," the management plan, et cetera, are nice to have but may not really impact what it is this measure is addressing.

And so to the extent – this is really more of I guess a staff question ultimately and then a committee discussion. But you know, several of these measures really are more sensitive to really being impacted by non-medical interventions.

And as we see the accountable care organizations and as we see other, well, hospitals of community benefit, et cetera, emerge and possibly focusing more in a population approach, they do address more than the medical clinical pieces.

And so is it appropriate for us to be addressing measures that impact hospital admissions but are impacted by non-healthcare types of interventions that really aren't specified anywhere in the background documentation for the measure.

So I agree that the measure, I think it's an important measure and I agree that it's something that can be measured, should be measured. But does the – do the developers need to provide more information on the ways to impact the measure in non-medical ways.

Female: (Ron) this is …

Thomas McInerny: I think we're talking about social determinants and Pell and you know, increasingly, not-for-hospitals are being required to address some of the community factors of the result of deterrence itself and would lead to decreased hospitalization.

I think this is a very, right now, as consumer's infancy and we're not seeing educational exception, I don't think we're seeing a lot of that planned activities yet. But certainly, it is an activity that accountable care organization hopefully will promote hospital's working more on the social determinants to tell to reduce hospitalization.

Elisa Munthali: This is Elisa from NQF and this is really good discussion and Ron, I think you remember from our past population health project. These issues of accountability, shared accountability and that locus of accountability came up particularly within the context of NQF. As, you know, the majority of the measures that we've seen have been focused on the clinical care delivery system. But those are not the only measures that have come through and what we would like to encourage the committee to do is to continue this discuss but to look at the measure as it s currently specified.

And, you know, this issue should come in as you evaluating. We should give consideration. But you should be evaluating against NQF criterion of important to measure and report. The science of the measure whether this information can be repeated or whether we're indeed measuring what was intended to be measured, the use and usability of the measure and the feasibility of the measure. So we just want to remind everyone that as you're having this discussion they may not look the typical measures that have come through NQF traditionally. But we should be looking at the measure at this currently specified and the important discussions to have.

Ron Bialek: I appreciate those sets of reminders and that's really use in the feasibility issue that I believe is what crops into this. So it's a valid measure, the way it impact though is in the use – and feasibility discussion.

Elisa Munthali: What if we get it if – if you look it from another standpoint which is the measure of itself will ultimately be of the measure of how well all those other things impacted it. So looking to see if we're able to fix some of the social determinant of how which – which probably for asthma more than almost any of the others we've looked at based on this group are affected by the environmental component. But ed result is how well you're doing is either missed school days or is hospitalization.

So whatever you're doing is going to show up an impact, this ultimate measure (inaudible) looking at and how I was hearing that conversation.

Adeela Khan: Are there others from the committee that would like to weigh in? We'd also like to get Patrick and Pam's thoughts on the discussion as well.

Eric France: This is Eric. I was thinking about how at Kaiser Permanente we have and another health plan. They have robust program to engage families with young children who have asthma to be sure they're on in health steroid to keep them out of the hospital and understand their environment settings and so, there is a lot of ambulatory care in communities that's focused on children with persistent asthma to keep them out of the hospital in spite of the environment triggers that are there. We're still working with him with that end in mind. So I think this is a strong measure in that regard.

Adeela Khan: Patrick or Pam, do you have anything to add?

Pamela Owens: Patrick, I'll let you go first.

Patrick Romano: Sure. This is a really terrific discussion and I think it's very similar to discussions that we've had with the clinical advisory panels for AHRQ. These measures and this one is really the prototype for these measures are intended as multi bacterial measures of population health and wellness. There's no question at there are social determinants as there are social determinants of –

of so much in health. We agree certainly that secondhand smoke exposure, air pollution, other environmental exposures are likely to be confounders.

We would be concerned about that if we were applying the measures to evaluate the performance of individual providers. But since the measure is intended to really assess what's going on at the community level, at the populations level, we think it's important for people to understand when and where these hospitalization rates are high and if they they're high partially because of environmental factors. There may be opportunities for public health intervention focus on those environmental factors reducing secondhand smoke exposure for example.

So I think we would agree with everything that's been said here and just again to emphasis that these measures do have to be interpreted in the context of population health. And certainly, some Medicaid Managed Care Plans have demonstrated the ability to reduce hospitalization rate even within high risk population base on socioeconomic factors.

Pamela Owens: And I think the only other thing that I would add to what Patrick is saying, I completely agree with what committee is saying as well as with Patrick is that perhaps on the RQI site. One of the things that we can do to enhance the understanding of the interpretation of these indicators and potential misuse to prevent the potential misuse of the indicators is to really beef up the documentation around the public health aspects of it and social determents of health which I think are critical.

The other aspect that I come back to particularly on this indicator is that if a child is hospitalized for asthma. That's different than if a child goes to the emergency department for asthma. They have to be pretty severe and a lot of things in the system would have had to have fallen apart for them to have been hospitalized for it. It's just interesting to me, especially if we grapple with on the RQI site we're developing prevention quality indicators for the emergency department.

I didn't bring that up in some of our previous discussions. But it one of those conditions that I grapple with in terms of preventability and where in the systems is it preventable.

Thomas McInerny: Agree.

Adeela Khan: OK, why don't we move on to the next measure? The last measure we have is 2372, that's breast cancer screening. This measure was developed by NCQA and I believe we have our developers on the phone. The lead discussants for these are Katie Sellers and Amir Qaseem.

Amir Qaseem: You want me to just kick it off first?

Katie Sellers: Sure, that'd be great.

Amir Qaseem: Sure. This is I think is the measure. It's been update of the measure because I think NCQA used to have a measure last year for year and a half ago at least that was the percentage of them in between 40 to 69, that was an NQF endorsed measure, which I think that also was taken away. It might have been because of the updated guidelines that have been coming up. So what this topic is addressing is the percentage of women 50 to 74 who had a mammogram to screen for breast cancer.

The numerator includes all women who received a mammogram, the denominator statement is between 50 to 74 for the December 31st of the measurement here. There's some denominator exclusion that pretty much a goal more in the progression of some clinical history not based on patient preferences in any way. If you look into the evidence, performance gap and priority items, I need to really go into detail. There's enough rationale that's been out there regarding the need for this measure.

The NCQA folks actually provided some that where – where is the performance course as well what's been going on. They provide evidence behind this guideline, major evidence behind this guideline that was presented I believe with the US Preventative Services Task Force guideline that came out. We actually gave it a grade B recommendation that we should be screening all women in age 50 to 74.

In terms of that priority, anybody has high priority measure. It says epi data and all that. That was provided. It was I think – the measure is well. We did a very good job with that.

Reliability and validity, again, I have not really any concern but there was some – I get a little confused when it comes to the measurement in terms of the time periods of the data. But I think that was previously endorse as well as the – I remember discussing that I don't really have any specific concerns regarding that when it comes to reliability and validity. And in terms of the data sources I think the information for this measure can be collected quite easily when it comes to various claims like the process measure. Of course, it's not related to any of the outcome.

I did not see any concerns when it comes to that as well as of course the usability and the details that go into this measure. My bigger concern when it comes to this measure is sort of a philosophical thing that whether we should be (inaudible) performance measures where the quality of evidence.

Something that I started with I think earlier this conference call as well as should we have measure where the quality of evidence has been classified as moderate. So if you look at the US Preventive Services Guidelines, they say the most of the evidence. And again, I'm not questioning any of that stuff. I mean the – this is not between 40 to 49. So I think the data is quite clear when it comes to between 50 to 74.

But even the taskforces, the most data comes for women in 60 to 70, so that's one thing. But of course, you can expand some of it in 70 to 74 and again, I'm not going to get into the reasoning behind the task force guidelines and all that. But still the taskforce gave it a moderate quality evidence and it is a great B recommendation. They are not saying they should be done in all women.

So this essentially goes back, they should be the – well, I think performance measures that are not based on high quality evidence because there's enough out there. I think that we should be doing where we clearly know there is

benefit, where we clearly know there is where we need to be improving clinical outcomes and where we have no doubt that there is no controversy that physicians should be doing a good job versus the measure where – again, that even the taskforce after evaluation of evidence is not giving it as a grade A, been given as a grade B.

So that's sort of a summary and again, of course, I can go into the reasoning why the task force gave that recommendation. That might of course help make certain decision. I don't think we're going to talk in too much detail about this measure today. That's actually from my end.

Katie Sellers: This is Katie. I was very much on the same page with Amir on this. I felt like everything was really very straightforward. But I did have a similar question about the task force and, you know, first of all, my question was is a recommendation from the preventive services task force something we need to dive into and look into or do we just say that's evidence enough?

And then second, how do you deal with a grade of a B as opposed to an A? So, I don't know if the staff can provide any clarification on those issues. That would be really helpful.

Amir Qaseem: One more thing I wanted to add and then we can, I think, have some of the discussions there. I think I got bothered in the way that currently the measure it. It might be OK to go with this measure if we can somehow fix the exclusion criteria because the exclusion criteria does not talk about anything about documentation of patient reasons for declining a mammography. I think it's quite critical because again, it's a grade B recommendation.

This was quite many times and that means there is a possibility there is a 30 percent of women who might say no. And in that case, I think that we need to have some sort of a modification to that.

Jacqueline Moline: This is Jacque. Just from my reading of the USPSTF guidelines, every single cancer screening test that they have approved is a grade B, there are a very few great A aside from like childhood vaccination and a couple of cardiac tests or basically checking blood pressure every five years. So every single cancer screening test is a grade B and they always have been. So, you

know, I think that needs to be context, it isn't that mammography is any different from colon cancer or cervical cancer.

And in terms of the age, I mean, while there's a discussion now, there's even in the past month, there's been discussion about whether it's valid over – the screening should be over age 70 but the USPSTF is going to lag behind in any kind of recommendation for a couple of years.

Male: So they are actually open now (inaudible).

Jacqueline Moline: They are looking at it now.

Male: Yes. The last recommendation as Amir said was moderate, certainty of moderate benefit and that was in 2009 and so now they're starting to work to determine what are the key questions that are brought forward, so I too have the same issues that Amir did about philosophically. This is a metric in general that's been flat for 10 years and declining across states in part because more and more women are saying you know, I see the value and benefits versus harms and I'm not particularly interested in being screened.

And it gives me a lot of pain and aggravation with my region as people are saying why aren't your the mammography rates going up higher and the primary care physicians are saying "Listen, I'm tried of asking a woman 16 times because of these metric out there that I need to perform, she said no. And so, we're moving more to metrics where the quality of the discussion is what's in the numerator rather than whether or not to get screened. And I would imagine over the next first year they'll just be more of that kind of general view that the rates aren't going to be getting better. It's going to be more about shared decision making with women about whether or not they wanted to be screened.

And the – I'm sorry. The impact is pretty low to a 15 percent reduction in mortality for the 50 to 69 year olds. So it's not like it's a great screening test with great importance in the population to be followed.

Adeela Khan: So, this is Adeela. We've actually just pulled up the NQF algorithm and you can see here that even if the grade is still B, it's still rated as moderate. So it

still falls, you know, as acceptable under the NQF criteria. It's just a reminder to put up their pretty reference.

Female: Perfect. Thank you.

Is there any other committee members who would like to say anything about the measure?

Marcel Salive: I think it was well articulated. This is Marcel. I think the key points were made about moderate evidence and moderate benefits. It's true, I would agree, the other comment about many – all the cancer screenings only have moderate benefit. So it doesn't make them good or bad. There's only so much you can do really.

Male: I think that's again, my point is again, there's few good clinical practices, there's a few things we should all be doing, but does every single good clinical practice needs to translate into a performance measure and that's where I think I really start struggling sometimes and I'm OK with it just like a high quality evidence behind something and yes, we should all be doing this. And we can see there's a performance gap over there.

Where is moderate especially, where is a lot of value judgment, some of you were saying. It's a lot of women wants to discuss, we're doing this. I think there are bigger things we can – more important measures that are out there we first need to fix before we start getting into some of this middle ground.

Adeela Khan: Does anyone from NCQA who would like to respond?

Sepheen Byron: Hi. This is Sepheen Byron, Mary Barton and Stephanie Rodriguez from NCQA.

Mary Barton: We appreciate being given the opportunity to respond. You know, I think we are so glad when we have a US Preventative Services Task Force recommendation on which it is based the measure because I think notwithstanding this between A and B and, you know, moderate benefit or high benefit. There are very few NQF endorsed measures that would need this standard of evidence honestly. And so if you want to go and sort

readdress the entire endorsement enterprise I would support and participate in that. But I don't think that that's necessarily where the breast cancer screening measure would fall down in all honesty.

But that being said, we appreciate the careful discussion that is measured is getting in the workgroup and I look forward to the continued discussion in the steering committee.

Amir Qaseem: Mary, this is Amir. First, thanks so much for responding. If the – my only point is this is a colorectal cancer screening, colorectal cancer task force today. So I'm not arguing about that but my argument is the B recommendation. So if you're going to account or even if you're going to look at the task force, I think it's OK to take taskforce A screening recommendation but the taskforce is not even sure about this one.

Mary Barton: Well, I don't think that's quite fair. Is not to say that the taskforce is not sure. I think that what the taskforce and you and I can discuss this maybe applying. But, you know, the – what the taskforce has been carefully trying to separate the impact on a population and when you have a moderate benefit, that's the someone quoted before the 15 percent reduction in mortality for people between ages 50 and 69.

They, you know, their algorithm for assigning a letter grade, you know, really puts them towards into the B category. The only way to get in the A category is to have high certainly about a high met benefit.

And as was mentioned before, things that impact the lives of young people are, you know, situations where we can see a very high net benefit. But it's hard in the case of saving a few years in those later life to see a high net benefit.

And so I think that's what Marcel was saying about the cancer, the field of cancer in general. But that being said, I think that the point about performance measurement here and you know, for health plan, just to bring up one of your former points, typically health plan measures, we do not allow space for patient refusal because it's not feasible to collect and in fact it's very,

very consistent to imagine that comparable entities will have comparable rates of refusal.

And in general, when we've investigated this, we find that it's really below 5 percent that you would actually find that happening.

Now, I'm not saying that that might not change over time and I certainly can appreciate the point that was made about in the future wanting to use measures that potentially use electronic health record data that ascertains the quality of the discussion, the quality of the decision making.

But we're not yet at that point I would say in the performance measure world.

Arjun Venkatesh: It's Arjun here. Can I build on that point for one second? I think that that's a great point about when we think about these measures, particularly that this being community level measures that we need to focus on things that we believe where there is community level differences between communities that are driven by community level interventions.

And in the case of refusal, I could believe the kind of premise that there may not be a systematic mal-distribution of refusal between counties. And the reason I think that's important is because when I look at this measure, for me where I think the opportunity for the measure have improved outcomes is in thinking about communities where they may be larger disparities and I live in one of these in New Haven. We're about 20 percent uninsured, 25 percent on Medicaid. And I would imagine that our county probably has a lower performance rate on this and there are a variety of community level intervention we're doing to increase mammography rates particularly among lower income or Black or Hispanic women here.

And so to me this measure is one that I could see having value from a public health perspective because we sit on the mammography spectrum probably on the underuse aspect.

And a lot of I think from where the concerns are and where performances are improved and where a lot of this comes from, it's from the side where there is relatively higher performance and people are at a point where they at a

community level may be hit higher rates and they're at this point where they're getting more refusals and that's the people that aren't not getting screened.

And so what I'm wondering is, you know, if we really think that some of these things are fairly well distributed across counties, does this measure still have value in terms of reducing disparity?

And to me, I think to some degree that's the case but I don't know this evidence on. I'd appreciate maybe some more evidence around that actually county level or however, whatever the denominator for this in terms of community level differences, probably it's county for me where these disparities will be revealed and I think some data there would help.

Sepheen Byron: Thank you. This is Sepheen Byron from NCQA and I do want to note. We do get information, geographic information. It's at the health and human services definition of a region so it's more of a regional level.

And when we look at the data by region, you do see some differences in the rates. The more pronounced differences we do see are between the commercial and Medicare and Medicaid, plans with Medicaid rates that tend to lag more.

And so, and then you know, additionally we'll look at the different percentiles of the performance rate, and so yet you do see quite a difference between the 10th and the 90th percentile plan. So we do see enough of a variation where we believe there is still a gap in care.

Marcel Salive: This is Marcel again. I wanted just to comment on the level of evidence and I think, you know, one of the luxuries we do have here is that the task force, US Preventive Services Task Force, you know, only really rates effective services A or B. And so we have one that's very effective but among those effective services, they never really rank anything, they are just commenting on the evidence and the amount of benefits.

There was a Prevention Priorities Task Force that was separate from that other task force and they looked at it for mammography and it was actually pretty well ranked for the, you know, the potential effectiveness plus cost

effectiveness basically. And so this is a very good service. I think we had a very good discussion on that.

Eric France: Yes, I think – this is Eric again from Kaiser. I think that analysis certainly didn't list breast cancer as one of the top opportunities for prevention and interventions across the country. It was lower to the middle of the pack as I remember it.

And there is this variation that is probably more than just access, the intermountain states typically are 5 percentage points lower than the coast. Northeast is particularly high in screening rates and Mid-Atlantic states as well.

And some of these I think are the different priorities and views of people who have similar access to healthcare in these communities but still have different rates based on where they live in the country.

And I can imagine given what we know today, given where this measure is today that this would continue as an endorsed measure by the committee. And I wouldn't be surprised if in this next review by USPSTF, it may even drop from its B status to something else. But that's for a future time for this committee to review.

Adeela Khan: Were there any other points anyone wanted to bring up about the other criteria, scientific acceptability, or usability or feasibility?

Ron Bialek: This is Ron. I just have one question which is about the disparities issue here and again, this is a single measure for the population and/or for segments in the population.

And if there is improvement within the plan, we don't necessarily – I mean the improvement could be far greater for one segment of the population and lower for another that's not necessarily teased out here.

Is it the advice of staff to keep these measures as a single measure or where there are data available that do allow you to tease out income in particular, that there would be a measure that might also be related to an income level?

Elisa Munthali: Ron, sorry, this is Elisa. Could you explain a little bit …

(Crosstalk)

Ron Bialek: I'm sorry. I was just trying to confuse you. I did a good job.

Elisa Munthali: Yes.

Ron Bialek: Right. I mean we know here it was pointed out with the Medicaid plans, the utilization of mammography screening is far lower. And even in non-Medicaid plans, typically you find the lower income populations having lower screening rates.

And so the question I have is, in all of the measures that we've discussed, there is a single measure and there could be a benefit for there being two measures around the same particular item like mammography screening and one that may be addressing the full population, another one that may be addressing the lower income population.

And I didn't know if that's just something that NQF shies away from doing or if it's inappropriate to do. I'm thinking about at the end of the day, you don't necessarily know if you're making an impact with regard to disparities or if you're really, you could be widening disparities and having a far greater impact for individuals at a higher income.

And so, you know, the data would show the rate increasing for mammography but in reality, the disparity could be getting greater.

Elisa Munthali: Yes. So Ron, that's a good question. Right now we actually have a project on risk adjustment and socioeconomic status that's going on. And they're hashing out a lot of these issues and whether or not measures should, you know, developers should stratify or they should risk adjust based on these issues that you brought up.

And you know, we don't – if a measure comes to us as you have suggested, a developer would put it in 1B performance gap to kind of point out those issues

of disparity where there is inequity in healthcare or in health that we would like to look at.

They may also specify their measure where they are risk adjusting but, you know, some have opinions about what that may mean or stratify so we can actually see the breakdown per income or per race and ethnicity or gender or gender identity and things like that.

Ron Bialek: OK, thank you.

Adeela Khan: Any other comments?

OK, (Amy), can we open up the line please for public comments?

Operator: At this time, if you would like to ask a question or have a comment, please press star 1 on your telephone keypad.

We'll pause for just a moment to compile the Q&A roster.

There are no public comments or questions at this time.

Adeela Khan: OK, great. So as far as next step goes, we'll be compiling everything that was discussed today. We'll probably have that out to the committee about two weeks before the in-person meeting. So that will be a compilation of all of the workgroups 1, 2 and 3.

In terms of, I know we've talked a lot about evidence and importance and how it relates back to this population health, social determinant measures. So what I've actually done is on the webinar you can see that I've added a link there for measure evaluation, criteria and additional guidance.

And I can actually have Elisa talk about that but it's essentially additional guidance for population level measures that the previous population level, the previous population health committee came up with. So Elisa, if you wanted to add anything about that.

Elisa Munthali: Yes, just simply this is something we discussed during the orientation call. And as I have alluded to before and I think Ron and Jason and others who are part of our first project, that was our inaugural project on population health.

And so, we were trying to figure out what space we were in within NQF and then evaluating those measures that might be to the social determinants. And so, what the committee did is to review our current measure evaluation criteria to see if those were applicable to population and community level measures. And so by and large, they agree that they were but there were some considerations and additional guidance that we needed to give developers and the steering committee as they were reviewing population and community level measures.

And so, it's not just nomenclature changes that we've made to the – that we've included but we've also included additional guidance.

So what you have in the left column is our measure evaluation criteria and what you have in the right column is what our committee came up with, the population health, measure evaluation, additional guidance and context.

And, Adeela, if you scroll down, we have made changes or expanded upon the standard guidance for population health in red. So what you see in red is specific guidance for population and community level measures.

We did not change the criteria. We just added additional guidance to help developers and to help the steering committee as we were engaging in this measure evaluation process.

So any questions that you have please send them to us but we thought this might be helpful for you as you are looking at the measures and for developers as well so you know how the committee will be assessing your measures.

Adeela Khan: And this has also been posted to the committee's SharePoint page as well. I just posted it so it's available for you to download off the web if you aren't able to get it off the webinar.

So our next workgroup call is actually on – It's on March 31st and we're going to be going over diabetes PQI measures. So feel free to call in if you're available. This was a really great discussion. I know a lot of you are new to the NQF and new to this whole committee process but really, this was a really excellent discussion.

And thank you all for your help. One last thing I would add is that if you haven't done your survey, then go ahead and turn them in before – as soon as possible essentially. We'd like to start compiling those.

Eric France: And can you clarify again for me, it's Eric. Do you want us to do that for each of the measures we reviewed today or the one that we are presenting?

Adeela Khan: We'd like it to be for the one that you are presenting but if you want to do it for all of them, you are more than welcome to.

Eric France: OK, thank you.

Adeela Khan: Yes. And also, we really like to thank the developers for being on the call today. This was a really great discussion. And with that, I will give you back 12 minutes to your day.

Eric France: Thank you, bye-bye.

Thomas McInerny: Thank you.

(Crosstalk)

Female: Bye.

Female: Bye.

Female: Bye.

Male: Bye.

END