

NATIONAL QUALITY FORUM

Moderator: Taroon Amin
April 3, 2013
3:00 p.m. ET

Suzanne Theberge: Good afternoon, everybody, and welcome to the first GI/GU Steering Committee conference call to evaluate measures for stage two of this project.

This is Suzanne Theberge. I'm the project manager. I've spoken with most of you via e-mail or on the phone. And welcome this afternoon.

If you have any trouble, technical issues getting in to the voice portion of the call, please send us a chat message using the chat box on the webinar and we'll do what we can to get you pulled in.

At this time, I'm going to turn this over to NQF's general counsel, Ann Hammersmith, to do our conflict of interest.

Ann Hammersmith: Hi, everyone. I'm going to give you a brief summary of the conflict of interest rules of the road. And, then, I will call you by name to tell us if you have anything you want to disclose.

All of you should have received a rather detailed form from us where we ask you certain questions about your professional activities, any consulting activities, research activities. For the purposes of this call, we ask you to disclose anything that you think is relevant to what's before the committee today. So, in other words, please summarize your resume. You don't have to tell us everything you've ever done. Just what may be relevant to what's going to be discussed today.

I also want to remind you that you sit on the committee as individuals. You do not represent your employer or anyone who may have nominated you for service on this committee. In addition, things that need to be disclosed may not be purely financial.

Sometimes, I hear people say, "I have no financial conflict of interest," which is great. But, you can have something that is a conflict or that it would be a good idea for you to be – to reveal to us even if no money has changed hand. For example, voluntary service on a committee where the committee's work was relevant to what we are doing here.

So, with that, I'm going to call your name, and tell us if you have anything to disclose.

Andrew Baskin?

Andrew Baskin: Yes. Hi. Yes. I'd like to disclose that one of the measure stewards for two of the measured (states) is Active Health, which is a subsidiary of Aetna where I work. And, because of that, I will not be voting and I will not be partaking – I will not be taking part in the discussion either.

Ann Hammersmith: OK. Great. Thank you very much.

Christopher Saigal?

Christopher Saigal: No conflict.

Ann Hammersmith: OK. If I mangle any of your names, I apologize in advance.

Liliana Bordeianou? Is Liliana Bordeianou on the phone?

Liliana Bordeianou: Hello.

Ann Hammersmith: Yes.

Liliana Bordeianou: Sorry. I pushed the wrong button.

Ann Hammersmith: That's OK.

Liliana Bordeianou: I figured out that it was me that will call when you started saying about mangling the name.

Ann Hammersmith: I hope that I didn't do too bad.

Liliana Bordeianou: No, it's fine. Liliana Bordeianou. I don't have any relevant disclosures. I did consult for American Medical Systems. But, I'm not constantly consulting for them. And none of the measures today have any relevance.

Ann Hammersmith: OK. Thank you very much.

Zahid Butt?

Zahid Butt: Hi. I'm a practicing gastroenterologist and CEO of Medisolv. We implement NQF-endorsed quality measures for hospitals and health systems. I also chair HIMSS National Quality Forum Task Force, which does get involved in some policy issues as it relates to eMeasures. I don't have any specific disclosures of any conflicts for this project.

Ann Hammersmith: OK. Thank you.

Robert Ellis?

Robert Ellis: I have nothing to disclose.

Ann Hammersmith: Thank you.

Nancy Faller?

Nancy Faller: I am a reviewer for a journal that covers clinical incontinence.

Ann Hammersmith: OK. Thank you.

Edward Gill?

Edward Gill: I have no disclosures.

Ann Hammersmith: OK. Thank you.

Johannes Koch? OK.

Jenifer Lightdale?

Jenifer Lightdale: Hi. I have no disclosures.

Ann Hammersmith: Thank you.

Richard Luetkemeyer? Is Richard Luetkemeyer on the phone?

Suzanne Theberge: (Natalie), we can see that Richard Luetkemeyer is logged into the webinar. Can you see if he is on the other portion of the line or something? If he dialed in?

Operator: He is not currently on the phone lines.

Suzanne Theberge: OK. Let's see if we can get in touch with him. Thank you.

Ann Hammersmith: OK.

Alayne Markland?

Alayne Markland: Just to be consistent, I do receive funding from the NIH and the Department of Internal Affairs related to the evaluation and treatment of urinary incontinence.

Ann Hammersmith: OK. Thank you.

John Morton? Is John Morton on the line?

Anne Pelletier-Cameron?

Anne Pelletier-Cameron: I have no relevant disclosures.

Ann Hammersmith: OK. Thank you.

Stuart Reynolds?

Stuart Reynolds: I have no disclosures.

Ann Hammersmith: Philip Schoenfeld?

Philip Schoenfeld: (No).

Ann Hammersmith: All right. Did I miss anyone? Do you have any questions about any of the disclosures that have been made today?

OK. Thank you.

Suzanne Theberge: All right. Thanks, everybody.

Now, just to introduce the project staff, this is Suzanne Theberge, the project manager, as I said earlier. And I am here with Reva Winkler, who is the senior director, and Evan Williamson, who is the project analyst for this stage of the project.

I know we have a number of measure developers on the line as well. So, I'm just going to call those up by organization. And if you could just let us know that you are on the line, that would be great. Thank you.

(Art)?

(Art): Yes. Here on the line.

(Jefferson Pattel): This is (Jefferson Pattel).

Suzanne Theberge: Great. Thank you.

NCQA?

Female 1: Hello. We're on the line.

Suzanne Theberge: Great. Thank you.

(Active Health)?

Female 2: Yes. We're on the line.

Suzanne Theberge: And (AMAPCPI)?

Female 3: Yes. We're on the line.

Suzanne Theberge: Great. Thank you.

And – OK. Now, I just – sorry about that. I jumped past the agenda. So, I just want to go over the agenda real quick for this call.

We'll be reviewing the measure evaluation criteria with Reva very briefly. And, then, we'll jump into the measure evaluation discussion, which is when we'll go over the measures that are on discussion in this phase. Finally, at the end, we will have an NQF member and public comment. And, then, we'll wrap up for the day.

So, with that, just a couple of housekeeping notes. As per usual on calls, please put your phone on mute if you are not speaking. Please do not put this call on hold because, then, your hold means that will play on the line.

And we have posted links to all of the submission forms along the side in the links box on the left-hand side of the screen so that if you would like to pull up any of the measure forms, they're right there if you don't already have them available.

And, with that, I will turn it over to Reva for the eMeasures evaluation overview.

Reva Winkler: Thank you, Suzanne. And thanks to everyone.

As we begin stage two of this project, we will be looking at the remaining criteria that NQF uses to evaluate measures. During stage one, you looked at the first criteria of important (information) report. We will not revisit those.

But, today, we will begin the last three criteria and we'll talk just a little bit about these. I know that we went over some on the orientation call. And, hopefully, you've had a chance to look at it yourself. But, this will be an opportunity to clarify any of the concepts or issues surrounding each of these criteria before we begin using them to evaluate the measures.

So, the other three criteria that we will be discussing in stage two with evaluations beginning today is scientific acceptability of the measures properties. And that's, essentially, reliability and validity and (that's some of the past criterion). The other criteria are usability and feasibility. And what – we're going to discuss the elements of those briefly as we go through them.

So scientific acceptability is the extent to which a measure, as specified – this is when we really look at the specifications for this specific measure whether it produces consistent and credible results about the quality of care when implemented. So, both reliability and validity must be – must pass.

In reliability, we are looking at precision of the specification. Are they well defined? Are the appropriate codes included? Is all – are all the details available such that anyone could implement the measure in a standardized uniform fashion so that comparisons could be made?

That's, essentially, the reason for the precise specifications. And, so, we are looking to you and your expertise to really look over those specifications to determine whether they provide all the information needed.

Reliability and empiric testing of reliability is required. There is a great deal flexibility on – over liability may be performed. It can be performed at either the data element level or the level of the measure score. However, in order to rate a high rating, it must be evaluated at both levels.

We'll talk a little bit more about reliability and validity testing in a moment.

Validity criterion – this is where we look at how the specifications relate to the evidence. And this is the evidence you discussed in stage one. And there is a very close relationship between the evidence and how the measure is specified and portrayed and whether they two are aligned. In order for the measure to be evidenced-based, it's important that the specifications do align with the evidence that you've already agreed meets the criteria.

Validity testing, again, is required, also, at either the data element or measure score level. However, we do accept phase validity, typically, at the measure

score level. However, if phase validity is the only assessment of validity, the highest rating you can give it is moderate.

We also look for justification of the exclusion. Are they evidence-based? Are they appropriate? And are all appropriate populations included in the measure?

Particularly, for (outcome) measures, whether there is a risk adjustment appropriate done and well characterized, we are looking for the ability of the measure to identify differences in performance. And, certainly, if the measure is presented from different data sources, the comparability and the equivalency of those results from those different data sources should be discussed. And, then, we are also looking for whether the measure is specified to be stratified for disparities.

So, reliability and validity criterion is a fairly large encompassing. You will vote independently for whether the measure passes reliability or passes validity.

Reliability and validity and very straightforward concepts. And, important, reliability is repeatability of precision. Validity is the correctness.

For those who like visuals, there are different combinations of reliability and validity. And we are looking for measure that are both reliable and valid. And I think this illustrates why very well.

Next slide.

Evaluation of testing. The empiric testing that is required is not particularly prescribed. And, so, it's important that you do look at the type of testing and the information provided. The specific attachment that describes testing is designed to take you through, in a logical fashion, how to look at the testing information provided.

And, so, an important question to ask yourself is whether the measure is tested at the level of the data element and/or at the measure score or both. And, so, it's – you can only rate a measure highest if it's been tested at both elements

or both level and, again, phase validity only if it's systematically assessed and the highest rating that can be granted is moderate.

Also, look at the evaluation of testing with the appropriate testing method used, considering the level of analysis, the type of measure, the topic, the potential sources of error. Are you convinced, based on the data presented, that this measure will produce repeatable, reliable and valid results?

And, then, was the scope of the testing adequate? Was there a large enough sample to be able to generalize the finding on reliability and validity to determine whether this measure, when broadly applied, is appropriately reliable and valid? And, then, were the results – statistical assessments of those testing are within acceptable norms?

So, those are the sorts of things we'd want you to look at. And those should be laid out very straightforwardly in the testing attachment.

When you're thinking about reliability, the threats to reliability are really an important aspect of your evaluation. The type of things to consider are ambiguous measure specifications, the lack of precision, the lack of definition, the lack of code set, the lack of the information necessary to be sure that there – the measure can be implemented in a standardized fashion.

When – the empiric testing – we do want to be sure that the testing is generalizable. So, consideration of the volume or the sample size, some issues around measures that – measure (rare) event can affect the precision or reliability in a measure score. It should be considered.

There are other possibilities of random errors, random errors in coding or transcription or random missing data. All of these things are potential threats to reliability that you should consider.

(Next).

The reliable is rated on a high, moderate or low or insufficient evidence or information. And, so, if you look at a high rating, it's the precise specification and empiric evidence of reliability at both data elements in the measure score.

Moderate is precise specification and empiric evidence at either data elements or measure score level. Low is, you know, issues with any of those – the empiric testing do not produce reliable results, there is ambiguity in the specifications or insufficient or where there is insufficient information.

And, so, those are the concepts around reliability. And I just want to pause for a second and see if anyone has any specific questions before we move to validity.

Do any of the committee members have any questions about what we mean for – about reliability and the rating scale used for this criterion?

OK. Hearing none, we will proceed.

Validity – again, are we getting accurate, correct, meaningful results? Threats to validity might be conceptual. A measure that's unreliable can, really, not be valid. Patients that are inappropriate excluded for measurement, such as a population that's not appropriate defined, differences in patient mix, scores that are generated from multiple data sources or methods where we don't know the comparability or the difference in accuracy or correctness or systematic missing or incorrect data, whether it's intentional or unintentional.

So, again, several potential threats do validity to think about when you are looking at the information provided in the testing. And ask yourself are you convinced, based on the information presented, that this measure will provide valid results when used in an accountability application and used to make comparisons?

Next.

Similar to the – our rating scale for reliability is the rating scale for validity. Again, high only if it's been empirically tested at both levels, the data elements and the measure score and the threats to validity have been assessed and addressed as well as the specifications are aligned with the evidence. So, you know, high is as good as it gets.

Moderate, frankly, is where we see most measures fall. The specifications are consistent with the evidence, testing is done at either the data element or measure score or there's been a systematic assessment of feasibility and threats to validity have been addressed.

Low is when we don't meet those criterion one way or another. And, then, insufficient information is when validity has not been appropriate assessed.

Any questions from anyone about what we mean by validity and the rating scale used to evaluate that criterion?

Christopher Saigal: Yes. This is Chris Saigal. It would be great if we have a cheat sheet for both. Is that possible?

Reva Winkler: How are you thinking?

Christopher Saigal: (Sorry) (inaudible). Maybe – I don't know if you could e-mail it to us or we could just sort of review briefly again before we vote to sort of make sure we (inaudible) (some of these rules).

Reva Winkler: We're thinking about how we might be able to do that.

Christopher Saigal: OK.

Reva Winkler: OK.

Christopher Saigal: Just simply show the slides when we're voting.

Reva Winkler: Well – yes. One of our issues is the technology we're going to use to collect votes. Can – I think we can go back to this slide. But, at the same time you're voting, we need to be on the voting slide to collect your vote.

Suzanne Theberge: Yes. We can't – we will be able to pull this up again. But, we can send you these slides so that you'll have them. So ...

Christopher Saigal: OK.

Suzanne Theberge: ... we'll get you that in a few minutes.

Christopher Saigal: Thank you.

(Shawn): And, Suzanne, if I may, I'm actually capturing that image now. And I can simply put that as a PDF in the links box for you.

Suzanne Theberge: That will be perfect. Thank you.

(Shawn): We're taking care of that now.

Christopher Saigal: Thank you.

Reva Winkler: That would be good. Can you do that for the reliability rating slide, too?

(Shawn): Let me see if I can grab that from the back end as well. We can, also – if you can show it briefly on the screen, I can grab it that way.

Suzanne Theberge: Sure, let me – here we go.

Richard Luetkemeyer: This is Ric Luetkemeyer. I don't know if you can hear me.

Reva Winkler: Great. Hi. Thank you for joining us.

Richard Luetkemeyer: Yes. I had trouble with my computer. I'm sorry.

Reva Winkler: No problem. Are you – everything working now?

Richard Luetkemeyer: Yes. Everything – I'm on the cell phone. And my computer is working. But, I can't communicate through it. (It's the) computer.

Reva Winkler: OK.

Richard Luetkemeyer: I have nothing. I have no disclosures or a conflict of interest on anything we're discussing today.

Reva Winkler: OK. Great. Thanks so much. I was going to ask you for it. (But, you beat me to it).

All right. (Shawn), do you have the reliability rating scale captured?

(Shawn): I have it captured. And if you give me about 20 seconds, I'll grab leadership in a second and load that into the links box.

Reva Winkler: OK. Fine. That is fine. But, we'll be able to continue on if there is any – any other questions about validity – the criterion on validity and the rating scale?

OK. Then, let's move on to the next criterion, which is feasibility. And that's the extent to which the required data are readily available, retrievable without undue burden and can be implemented. And, so, the real crux of it centers around is the data generated during the care process speaks to burden of data collection? Is the measure available in electronic sources? Again, burden of data collection.

The data collection strategy can be implemented – has that been demonstrated either through use or through testing? And, so, we are talking about things like the type of data – the data source, the burden on data collection and whether we know the measure can be implemented.

So, are there any questions about the criterion of feasibility?

OK. The fourth criterion that was – we want to talk about is usability and use. And this is the extent to which potential audiences – and we want to think very broadly – including consumers, purchasers, providers, policy makers, really just about anybody that has an interest are using or could use the performance results for both some type of accountability application or performance improvement activity to achieve the goal of improved performance and high quality healthcare.

So, the sub-criterion within use and usability is is this measure being used for some accountability or transparency application? Accountability are types of applications that are things like pay per performance, licensing, accreditation. And the transparency would be more like public reporting (but) other types of external uses of the data. These are high-priority uses of performance measures for NQF.

The second part of the criteria asks about what do we know about feasibility – of the ability of this measure to drive improvement? What sort of progress

towards achieving the goal of perform – improved performance do we know, particularly for measures that have been endorsed for a while or used for a while, trend data, data over time, seeing whether monitoring performance results in improvement and – or not. For new measure that might be a little bit more difficult to determine, but certainly for maintenance measures, it would be something important to consider.

And then, again, benefits outweigh the harms. This is the – where we consider the unintended consequences for the use of a measure to either individual or population. Consideration of potential unintended consequences is an important aspect of it.

Certainly, measures that have been in use for a while where there has been feedback and better understand of how well it works in the fields, they would likely to have some information around potential unintended consequences. Newer measures, probably less so, though a consideration of that is an important thing to consider.

Are there any questions from the committee about the criterion of use and usability?

Great. Both feasibility and use and usability are rated on a high, moderate or low rating scale. Again, these are qualitative assessments on your part to the degree that they need these sub-criteria.

(Meet) – unlike reliability and validity, which are must-pass criteria, feasibility and usability are not absolutely required. But, you will factor your ratings on – of usability and feasibility into your final recommendation vote on whether the measure should go forward for endorsement.

And, then, the last thing we want to talk about ...

Suzanne Theberge: Next slide, please.

Reva Winkler: Thanks – is whether we have any related or competing measures. Sometimes, we see measures from different developers or from different settings of care that, essentially, address the same process of care, perhaps a different

population, perhaps inpatient versus outpatient. But, regardless, the measure construct should be harmonized to the degree possible within the data system.

And if there are measures that are very, very similar, addressing similar populations, having multiple measures is not particularly helpful to the field. And, so, we will want to look at the competing measure side by side to determine how best to either merge them into one measure or be sure that we have the best measure recommended for endorsement among groups of similar and competing measures.

So, those are sort of the basic outline of the criteria that you are going to be looking at today. We're going to ask for individual votes on reliability, validity, feasibility, usability and use. And, then, a final recommendation on whether the measure should go forward for endorsement.

So, does anybody have any questions about the evaluation process that we will use today?

OK. So, I'm going to turn it back to Suzanne. We're going to talk about how we're going to collect your votes.

Suzanne Theberge: OK. Great.

Before I talk about that, it looks like (Com Partners) has managed to post the PDF of those voting scale. So, (Shawn), can you explain how folks can access those?

(Shawn): Absolutely.

If you would direct your attention to the links box to the left of the slide, you will find several links there that were referenced at the beginning of the call. And link number seven would be the NQF scale. That is a PDF document. If you click on that link, it will open a separate Web browser window and will not interfere with your viewing of the slide and the questions for voting.

So, it will open a separate Web browser window. You will be able to keep that PDF document open. Both scales are on the same page.

Richard Luetkemeyer: Great. Thank you.

Suzanne Theberge: Great. Thank you.

OK. So, we are, as you know, going to be voting on the measures on this call. And measures need – we need to vote to decide whether or not things will be moving forward. So, we had set up some online voting in this webinar.

The committee members are the only folks who have voting access. So, folks – developers and other folks who are listening in on the call, you won't see the ability to vote. But, you will see the results once they come through.

So, once the questions appear, what we'd like you to do is click in the correct box for your vote. And your response will be collected. It will be anonymous. You can change your vote at any time up until we move to the next slide.

So, as long as the voting slide is up, you can change your vote by clicking in a new box. And, as I mentioned in your e-mail – in the e-mail I sent, you need to have Flash installed for this to work.

So, we're going to test this out. We have a poll questions here. So, if the committee members could all try and vote, we'll see if this works.

I see that we have six of – 10, 12. All right. I think ...

Reva Winkler: Does everybody – do all of the committee members were able to vote? Does anybody have trouble registering a vote?

Christopher Saigal: I did. But, I want to make a comment. I voted red and (I'm moved to green and the vote got) registered.

Reva Winkler: Great. OK. So, you know for sure that yours worked. That's great.

Christopher Saigal: (Inaudible) that we (should be) voting twice. If you change your vote or you make an error, you should get to deal with that. Right?

Suzanne Theberge: It takes a – it takes a few seconds to be – to move itself from the count.

Christopher Saigal: OK. So, it will – it will switch over automatically?

Suzanne Theberge: Yes. It just takes a few seconds to load.

So, I see people are changing their votes right and left here. It looks like we've got 15 votes and we should have 14. So, one should disappear in a moment, I'm told.

But, anyway, it looks like everybody can vote. If you – if something happens and you need to leave the webinar portion of the call but you want to keep voting, let us know and we'll collect your votes verbally.

Reva Winkler: (Shawn), can you determine – we have 15 votes. But, I'm – my understanding is we only have 14 steering committee members who should be voting.

(Shawn): All right. Give me just one second. I can check on that in the back end.

(Zahid Butt): I think one of the developers is voting.

Suzanne Theberge: We do ask folks, if you happen to have voting privileges somehow but you're not a committee member, please don't vote because we'll have to go through (until you have all your votes later) and it will be challenging for us.

(Jenifer Lightdale): (I'm) (inaudible) your system.

Reva Winkler: (You know), that's fine. Well, you should be able to correct or change a vote and the old one go away.

(Jenifer Lightdale): OK.

(Shawn): And it looks like we're up-to-date now.

Suzanne Theberge: OK. Great. All right.

So, with that, it looks like everybody can vote. So, I'm going to turn this over to our co-chairs to begin the measure evaluation discussion.

Christopher Saigal: Great. So, I think the first one is (NCQA-1), which I'm going to read, I guess. And, then, I'm sure who's discussing for that.

Suzanne Theberge: Chris, we're having a little trouble hearing you. Could you speak up or ...

Christopher Saigal: Sure (inaudible).

Suzanne Theberge: ... move a little closer to your phone?

Christopher Saigal: Is that better now?

Suzanne Theberge: Yes. Thank you.

Christopher Saigal: OK. So, I was just saying that I'm not sure (who'll discuss it for the) NCQA measure.

Alayne Markland: Hi. It's Alayne Markland, I believe. Is that correct? This is me.

Christopher Saigal: OK (inaudible).

Suzanne Theberge: Yes. That's correct.

Christopher Saigal: So, Alayne, if you would – if you mind reading us through the (remaining three) criteria and your assessment (on how to meet that).

Alayne Markland: OK. Great.

Christopher Saigal: (Inaudible) measures, (too), (inaudible).

Alayne Markland: OK.

Reva Winkler: So, (this is really the look). It's easiest to do if we discuss the criteria in order and then you vote. So, we'll talk about reliability first. Then, you'll vote. We'll talk about validity. Then, you'll vote. And go through it that way.

Christopher Saigal: OK.

Alayne Markland: All right. Chris, do you want me to summarize this slide and kind of go through a description?

Christopher Saigal: Yes. That'd be great.

Alayne Markland: OK. This is measure 0098. And you can click on that full submission there. It's urinary incontinence, (assessment) characterization of plan of care for urinary incontinence women aged 65 years and older submitted by the NCQA as a maintenance measure. Is that correct?

Christopher Saigal: Right.

Alayne Markland: OK.

This measure assesses whether women 65 and older were provided appropriate treatment for urinary incontinence. And it, actually, has three different – not sub-scales but sub-ratings. Number one is for the assessment of urinary incontinence. And it's based on (numerous denominators) that includes the percentage of females aged 65 years and older who attended at least one qualified visit and were assessed during that visit for the presence or absence of urinary incontinence within 12 months.

The exclusion criteria for that includes the patients that were deemed either medically not necessary, that they did not need to assess it. And that can be seen in the actual document or medical (assessment) for the presence or absence of urinary incontinence within the last 12 month.

The second portion of this measure is characterizing urinary incontinence. So, it's taking all the women that have a diagnosis of urinary incontinence and having further information provided within the medical record on – for characters of their symptoms that had occurred at least once within the last 12 months in review.

And, then, the third piece of this is establishing a plan of care. So, the numerator and the denominator includes the percentage of females who have a diagnosis of urinary incontinence with a documented plan of care at least once within the last 12 months.

And the numerators there are for the assessment. So, it's just a documentation within the electronic health record or a paper chart for the presence or absence of urinary incontinence with a numerator of all women who visited an eligible provider in the measurement year.

For the second piece of this characterization, it's the number of women whose urinary incontinence was characterized of those that were actually diagnosed using numerator of all female patients with a diagnosis and visited an eligible provider (seeing) administrative data such as ICD-9 and CPT codes within the measurement year. And that's – that same denominator is also used for the plan of care measures. So, it's not only the characterization but the plan of care that uses the same denominator.

And, so, this measure type is a process measure. Its data sources uses administrative claims and paper medical record. And the level of analysis is the clinician or group practice or individual or clinical and team.

And, with that, I'm going to go ahead and stop at the basic summary and, also, the – in this – in this measure – and impact and report statement. The developers do mention the large impact of urinary incontinence, how common urinary incontinence and specifically in older women and the great impact on quality of life, not only from a medical standpoint but also a financial standpoint with citations given. And, then, they provide data from the PQRS, the Physician Quality Reporting System, that they have used to look at eligible provider level data that reported the number of providers participating (which has a) report on this quality measure.

And, with that, I'm going to go ahead and skip down to the point that we're going to focus on today, which is actually the reliability data first.

Are there any other questions before I present some of that reliability data?

Christopher Saigal: That's great.

Alayne Markland: OK.

And, so, if you scroll down to the end of this document and – the reliability data is presented – let me get the right page here on, I believe, page 31 of this document. And, from this, we not only need to look at the reliability data that it presents. But, we also need to consider the measure specifications.

Are they precise in terms of the numerator and the denominator? And what information is presented here for both of those? The data source is, actually, the Medicare Advantage Beneficiary. And, so, this is a very specific survey of those participants in Medicare Advantage Beneficiary program.

Are all the data elements, the numerator and the denominators clearly defined? And the denominators and numerators – (is that) presented there? And are appropriate codes included in this measure. And there does appear, from my perspective, to be appropriate codes for both ICD-9 and CPT codes to represent the denominators for the characterization and plan of care with the denominator for the assessment or are all women presenting to an eligible provider?

And, then, at what level was the measure tested? And, here – and please correct me if anyone (sees) – I believe that this level was just at the data element. So, we're not looking at a measurement score per se, although they do report the proportion of people – of their data that were tested for each of the different component in years 2009 and 2010. I see more of a data element in terms of the reliability data that they present here.

And the type of reliability testing, as I said, was the data element. And, in this table on page 31, when they had two different abstractors for an electronic health record versus paper record, the documentation process – they did find a high – a very high degree of reliability between two abstractors with a (kappa) value of 1.0 with a hundred percent agreement for the assessment of urinary incontinence based on the first measure, a (kappa) value of 0.94 for the characterization of urinary incontinence and a (kappa) value of 0.96 for the plan of care. And these, as I said, were using two intra-abstractor to look at how close their two data matched – their data abstraction matched in terms of reliability of being able to reproduce the same amount of data.

And, so, my interpretation of this result indicates that this meets criteria based on what we just heard from the slide set, that it is moderate reliability because the data element reliability is presented. And, then, you could say – and this can be for discussion – that they also present some data from one year to the next on terms – and, for me, it looks more of an improvement, not necessarily reliability data.

And I'm going to go ahead and stop there and open up to any other comments specifically about the reliability of this measure.

Christopher Saigal: We had the data elements (been tested) and not the measures in terms of reliability with the (kappa) testing?

Alayne Markland: That's what I – this is what – in my review of this proposal by the developers, I believe it is at the level of data element.

Christopher Saigal: That will be like the numerator or the denominator. But, the – on the – on page 31, (inaudible) (looking at that whole measure agreement) (inaudible) (numerator)?

Alayne Markland: Yes.

Christopher Saigal: Measure A (inaudible) and measure B (inaudible). So, (will the B) reliability of the measure?

Alayne Markland: It's not – and that's a good question. It's not a score. But, yes, it is reliability of the measure itself.

Christopher Saigal: OK. (So, this is) a little bit higher, (at the level of) (inaudible)?

Anne Pelletier-Cameron: (This is) Anne Cameron chiming in. Just if someone can, maybe, give me an explanation of the difference between the two because I think I'm a little bit confused now.

Christopher Saigal: I think – (well) (inaudible) was that if they're just looking out whether they are capturing a numerator value accurately – (so), the number of women who are intervened on with a plan of care for incontinence, that's one level of reliability. But, it has (the) actual like the (numerator) and the denominator

functions together in a – in a (task). That’s a higher level of reliability (according to the) (inaudible).

So, how (I’m looking at this page was that), you know, they were looking at the actual measure – (I’m on page 31 again) – and looking at the (intra-observer) agreement (inaudible) agreeing on whether that measurement (is met) or not.

Reva Winkler: Chris, this is Reva. Perhaps, your measure developer could clarify their intent?

Christopher Saigal: Sure. Are they available to answer our question?

Erin Giovannetti: Hi. Can you hear me?

Christopher Saigal: Yes.

Erin Giovannetti: OK. This is Erin Giovannetti from NCQA. I’m not quite sure how NQF differentiates the two. What was testing here was meeting the numerator criteria. And the reason that this is not focused on meeting the denominator criteria, which is having a qualified visit because we didn’t do an evaluation of medical records for people who didn’t have a qualified visit.

So, I don’t really know how that fits into the NQF framework. But, I would say that we call this data element because it is the – meeting the numerator. There is no score (out for this) measure. So, we couldn’t really do a (signal-to-noise), Chris.

Christopher Saigal: So, this is just a – this is (an agreement of finding the) numerator, then, not the whole measure.

Erin Giovannetti: Yes. Yes.

Christopher Saigal: OK. Then, (it’s a) lower level of rating (inaudible).

Reva Winkler: Chris, this is Reva. I don’t know that I would characterize it as a lower or higher level. They are just – they are different.

All right. And, so, what we require is evaluation at both levels to get a high rated – or rating or either level at a moderate.

Christopher Saigal: So, it's a moderate rating (inaudible).

Reva Winkler: All right.

Christopher Saigal: OK. That's helpful. Thank you.

Alayne Markland: Yes. Chris, that's – this is Alayne again. This is what I would (deduce) from both what the developer just said and from our discussion, that the data element portion is here but not both.

Christopher Saigal: Great. OK. That's helpful. Thank you.

Alayne Markland: And, so, just to summarize, in our – in the preliminary vote, this group voted four high, three medium, one low and zero indeterminate on reliability.

Stuart Reynolds: This is Stuart Reynolds. Am I on? Can you hear me?

Alayne Markland: Yes.

Stuart Reynolds: So, one of the questions that I have with this and that sort of bugs me a little bit is, in the part A, which is the assessment of urinary incontinence, even though they give this data which shows that when you use a couple of abstractors, they were pretty consistent in finding it, it's still not clear to me what the criteria are that means someone has been assessed or not. There is – there is not specifics in the description of what constitutes (the) minimum level to be an assessment.

And I'm concerned that that could, on one end, maybe by design, it's obviously general and not specific. But, I'm also concerned that that could be used in many different ways. And the example that I use is can (even a patient, when she comes into my clinic is) – you know, this review of systems (a) sheet and they can check off all different things.

And, so, you know, if this measure was looking at congestive heart failure, well, probably, a hundred percent of my patients have been assessed for CHS

because they checked the box yes or no. But, I don't know that that's exactly capturing what this is – intends to do. So, there is not quite enough detail, I think, in some of the description as they – exactly, what are the abstractors looking for that constitutes the assessment?

Alayne Markland: He does have a great comment. And I do agree. And I think (we see), you know, (inaudible) reliability here. And it's not as clear cut for that numerator. And I think this was also brought up before in a – (just in a comment) to our previous review, that they do realize that this is a – I don't want to speak for the developers, but that this is a potential problem and they're trying to clarify it with some further work on this. But, I do agree with your comment as well.

Erin Giovannetti: This is Erin, the measure developer. I don't know if you want me to comment now on this or not. I can wait.

Christopher Saigal: Sure. Sure. Go ahead. Go ahead. It's fine.

Erin Giovannetti: So, assessment is just what their documentation of yes or no urinary incontinence. It's not intended to be anything more than presence or absence of urinary incontinence. The characterization is, really, more of the formal workup of the urinary incontinence.

So, it's sort of considered the first step in clinical practices, like you said, a review of systems where you ask them once, "Have you had any urine leakage?" That would be the assessment. And if you ask everybody, then, you get a hundred percent.

The characterization would be the next step, which is that you, then, talk about the urine leakage and you figure out what are the sources of it and what type is it and what are some treatment options. So, when we say assessment for the presence or absence of urinary incontinence, that's really all we're looking for.

Alayne Markland: OK. Great. Thank you for clarifying.

Zahid Butt: So, all three of the components in the numerator have to be present for it to pass?

Alayne Markland: This is not a composite measure. It's not an all or nothing. This is a measure with three different indicators. So, you get three separates (ratings).

Zahid Butt: OK. Thank you.

Christopher Saigal: All right. Let's keep going.

Reva Winkler: Chris, this is Reva. I did have one question for clarification around the specifications. In terms of the title of the measure, it says it's an administrative measure. And I wondered what was meant by that.

And there seems to be an inconsistency (in some parts) of this submission where it talks about the data source. But, the data source that was tested was checked as abstraction from either paper or electronic health record. And, so, I think there may be a little confusion there ...

Christopher Saigal: Yes.

Reva Winkler: ... because the testing is only at the abstracted level.

Christopher Saigal: So, there's no administrative testing done.

Erin Giovannetti: So, Reva, this is Erin once again. This is because this uses CPT-2 codes which are generated through either abstraction of the medical record or marked during the – during practice. So, they're administrative in that they, you know, are a code. But, they are – so, they are looking for the presence or absence of a code.

The testing that was done was to determine (inaudible).

Suzanne Theberge: Getting some (feedback) on the line. Could you mute your line?

(Natalie), can you mute that line?

Operator: It's muted.

Suzanne Theberge: Thank you.

Christopher Saigal: Thanks.

Go on.

Erin Giovannetti: So, what the testing does is that it looks for a particular CPT-2 code. And, then, it uses two abstractors to see if they agree with each other about the (rating of the) CPT-2 code.

Reva Winkler: Yes.

Erin Giovannetti: So, if they – if the CPT-2 code says yes, this was assessed that the abstractors agree with each other that, yes, it was assessed.

Christopher Saigal: (Inaudible) presence or absence of code. (It wasn't) a chart review.

Erin Giovannetti: It was a chart review – to review the chart. And if the chart matched what was in the code and (inaudible) ...

Christopher Saigal: OK. OK. (Inaudible).

Erin Giovannetti: ... (inaudible) two reviewers and if they agree with the code.

Christopher Saigal: I got it.

Reva Winkler: All right. Good. Thanks. I don't (know the steps is all that clear and the data). But, that's good to know. Thank you.

Christopher Saigal: It is really – in practice, it is an administrative measure because it's essentially part of PQRS reporting for something (inaudible).

Reva Winkler: OK.

Christopher Saigal: OK. So, is that it for reliability and validity?

Reva Winkler: It's just for reliability, I think.

Christopher Saigal: OK.

Reva Winkler: We need to – so, if you're ready, we can move on to the voting slide.

Christopher Saigal: Yes. (Inaudible) any comments from anybody else (or) questions?

OK. I think we're ready for the (inaudible).

Reva Winkler: OK. Here's the voting slide.

Christopher Saigal: OK. So, (we've got good votes here). We've got three votes for moderate. (Inaudible) reliability (inaudible). (And we go to) 14. Right?

Suzanne Theberge: There should be 14 people voting. Maybe actually 13 – maybe 13 people voting right now.

Christopher Saigal: (I think we're missing) one vote.

Johannes Koch: OK. So, this is Johannes Koch. I got to the call a little late. So, I'm not exactly sure where we're voting. I've heard the last part of the discussion. But, I missed where we're voting. So, I apologize. That's probably me who is missing the vote.

Christopher Saigal: Johannes, are you on the Web site, the link that was provided in the e-mail?

Johannes Koch: Yes.

Christopher Saigal: And, then, (inaudible).

Johannes Koch: And I see – I see that – the slide right now that shows what the voting is. But, tell me where in there I'm ...

Christopher Saigal: Just click (on it), Johannes. It's interaction.

Johannes Koch: OK. Sorry.

Suzanne Theberge: You just to click on one of the boxes next to – yes. There we – yes.

Johannes Koch: OK. Sorry.

Suzanne Theberge: OK.

Christopher Saigal: (Inaudible) easy. (Inaudible).

Johannes Koch: I know. I didn't think it could be that easy. Sorry.

Christopher Saigal: All right. That's great.

So, (inaudible) to validity now?

Alayne Markland: All right. Thanks, everyone.

All right. If you – if you will take that same document and scroll again – we're on page 31 to page 32 – that – the developers do report some information on phase validity from this. And, so, the additional things here that I just want to go over in terms of what the validity testing are is that the only – that use an expert panel – what appears to be from 23 experts in the field – and there's a long list of names provided in the documentation.

And I'm assuming some of these were some of the expert panel and that in this validity for does this measure appear to, actually, measure what they believe it's going to measure, the expert panel, on a one to five scale for item A, the assessment of urinary incontinence, rated a – for a mean of 4.2 with 82.6 percent of those experts saying they either agreed or strongly agreed that it is – it accurately distinguish good and poor quality.

In item B, using the same sort of one to five rating scale from these 23 experts, they had a mean rating of 4.1 and 78.3 percent said they either agree or strongly agree that this measure can accurately distinguish good and poor quality based on the way the measure is designed.

And, then, for measuring if a plan of care was documented, the mean was 4.2 with 82.6 of the respondents either agreeing or strongly agreeing that the measure can accurately distinguish good and poor quality based on the plan of care. This assumes that the plan of care and other measures are looking at all the specifics that would be appropriate for urinary incontinence care and that the authors and developers do mention multiple types of care including both

the non-surgical behavioral changes, exercises and surgery as part of that plan of care.

And, so, from the level of data in the slide presented on what the interpretation, I believe the interpretation are results (are). It's that this phase validity was systematically assessed with this expert panel. And one of the limitations that this – the measure was not stratified by any disparity – and the developers do make a care that the data in the literature is conflicting on any health disparities in the presence and absence of urinary incontinence and (health-seeking) behavior and that they did not stratify according to disparities based on current data available in the literature on this.

And, so, for these reason, I would rate the validity at moderate based on the evidence provided – the phase validity evidence that's provided and the fact that the measure was not stratified for disparity. Excuse me.

Christopher Saigal: Great. So, basically, its phase validity is the – is the major way to measure validity that they have chosen. There are other ways to measure validity which they haven't approached. But, I believe in our orientation slide that phase validity was acceptable as a method. Is that right?

Erin Giovannetti: This is the measure developer. I'd like to make a comment when I – when I can.

Christopher Saigal: OK. But, phase validity is – was acceptable.

Reva Winkler: Yes, it is.

Christopher Saigal: Yes. So ...

Reva Winkler: (It is) acceptable.

Christopher Saigal: OK. And did you have a comment that you wanted to make?

Erin Giovannetti: Yes.

We, actually, also have critical data elements of validity. But, it counts for both reliability and validity. So, the data (above) is also evidence of validity as well as reliability.

Christopher Saigal: Yes. That's (inaudible) source (you're trying) to look at the chart data to see if it reflects what the actual code is indicating.

Erin Giovannetti: Right. We just didn't think it was necessary to repeat the same testing data over again in the validity. But, per the NQF evidence task force – I'm sorry – (inaudible) (testing) task force report, evidence of data element for reliability counts for both reliability and validity.

Christopher Saigal: Yes.

Alayne Markland: I'd also like to point out that at the bottom of page 33, the developers also present results based on the exclusion and they showed that from the two abstractors, the rate of exclusion, although (is) 90-percent agreement, the (kappa) value is 0.6 there for the exclusions. And they do note that there was a small number or a small sample size of the people that were actually excluded from that assessment.

And maybe the developer can talk about it. (It is) an overall rate of exclusion or it look like this is just for the assessment for urinary incontinence exception and some of the other exceptions were not included in terms of their exclusion criteria for the statistical data presented.

Is that correct?

Erin Giovannetti: Yes. There (was, then, the) sample testing that we did. We only found one exclusion. And it was before the assessment of urinary incontinence and it was such a small sample size and what's (why the) (kappa) is so low.

Zahid Butt: So – Chris, this is Zahid. Can I ask Reva, I guess, maybe a general question?

Reva Winkler: Sure.

Zahid Butt: So, Reva, I think when we have the opening slides, it was mentioned that if something is unreliable, it is invalid. So, I suppose the developers is making the contrary point, that if something is reliable, is it also automatically valid?

Reva Winkler: I don't you can't make that. But, at the data element level, empiric testing of reliability of the data elements, because you are comparing the reproducibility with the authoritative source, it does also address validity.

Zahid Butt: But not at the score level?

Reva Winkler: No necessarily. No.

Zahid Butt: OK. Thank you.

Christopher Saigal: OK. Any other comments regarding validity (inaudible)?

OK. I think we're ready to vote, then.

OK. So, well.

Suzanne Theberge: All right. I think that we have all our votes. And (I actually am) supposed – (we lost one now). We're at 13.

Christopher Saigal: Thirteen. Correct number, I think.

Suzanne Theberge: Great. All right. So, we have 11 moderate, 2 low.

Christopher Saigal: OK. Let's move on.

Alayne Markland: All right. Moving on to feasibility, some of the elements in feasibility include – is – are these required data elements routinely generated and used during care delivery? And I would say say. These data elements, in terms of the assessment and both characterization and plan of care are routinely generation and, potentially, should be (in the) plan of care delivery.

And are these elements available in electronic form and other electronic sources? And I would say yes, both from an administrative (point) of view and as well as other data sources.

And is the data collection strategy ready to be put in operational use? And I would say yes because this measure has already been used in the PQRS system with data reported on those that chose to participate in this measure in the PQRS system.

And, with that, I think we (might) – rating of feasibility would still fall in the high to moderate range. And I think I would – I would actually pick moderate for this. And as the comments on our preliminary testing and as, also, has been discussed, is that data element needed for the numerator there is still a concern. And, if used more widely, it may be more difficult. However, I still think this is feasible and would give it a moderate rating.

And I'll open it up for discussion where – on feasibility.

Christopher Saigal: (I think it's) a very good assessment. Anyone else?

OK. Let's vote.

Suzanne Theberge: All right. It looks like we need one more vote. There we go. Three high, 10 moderate.

I'll move on. OK.

Alayne Markland: All right. And the last thing, I believe, we are voting on is usability. And in terms of the usability, for maintenance measures criteria, this measure used in, at least, one accountability application and is this measure publicly reported? And, given the use on the PQRS system, I believe the answer is yes, going to the results presented in this – in this application from the developers and to the Web site.

And can this performance – the performance result use to further the goal of high-quality efficient healthcare? And the developers do present some data on performance. And I wanted to point those out that they are in – assess for performance is on page 34 of the document.

And as in the preliminary review, one of the reviewers did comment on the performance data did not well – did not do very well to discriminate above the

10 to 25 percentile. And that's specifically for that (option) – the first – the assessment piece. However, the characterization and plan of care look to – are doing better in terms of the performance and the scores across the measured entities.

And, so, I just wanted to show that data there for the three difference components of this measure. And developers do state, however, in the future, you know, that this initial data was from a self-selecting group and beginning in 2015, this may change based on changes in reporting measurement documents. And I do believe that this measure outweighs any potential unintended consequences. And I would give it a moderate overall rating – moderate to high for usability.

And I'd like to open up to any further comments.

Christopher Saigal: Yes. It looks like it's – I mean (inaudible) (are almost topped out). But – I mean, in terms of this (very useful measure). But, I guess, it's sort of saying that, essentially, there is a lot of (non-reporters). So, I guess, well – that, I think, is a reasonable – a reasonable explanation (to keep it in).

Any other comments?

OK. Then, we should vote.

Suzanne Theberge: OK. The voting is open.

Johannes Koch: A few more. So, I'm not getting it on mine here. This is Johannes.

Suzanne Theberge: You don't see the option to vote?

Johannes Koch: Not on this one. No. I don't know why.

Suzanne Theberge: It looks – you were able to vote earlier?

Johannes Koch: Yes.

Suzanne Theberge: All right. It looks – nobody else had trouble voting. Right?

Christopher Saigal: Nope.

(Reva Winkler): Is (Shawn) still on?

Suzanne Theberge: (Shawn), are you still on the line?

(Reva Winkler): Maybe not.

Suzanne Theberge: All right. Do you want – do you want to vote verbally on this one. And, going forward, we'll ...

Johannes Koch: Sure. It's moderate. Yes.

Suzanne Theberge: Moderate. OK. So, that's two high, eight moderate and three low. And let's see if you have trouble on the next one as well.

Christopher Saigal: OK. So, now, we're going to (take the overall vote) – (overall) (inaudible) for endorsement. (We'd vote) whether the measure meets the criteria for endorsement for this (inaudible).

Suzanne Theberge: I see 11 votes, 12 ...

Johannes Koch: (OK). (Give it a) moderate, again, for me, please. I'm sorry

Suzanne Theberge: It's a yes or no.

Johannes Koch: Yes. Sorry.

Suzanne Theberge: OK. So, you don't see the vote on this option as well?

Johannes Koch: No. I think I'm having a little (trolls like) Adobe Flash's. So, I apologize.

(Christopher Saigal): Yes. So, like – if you want to quit out of our browser and, then, open it back up and click the link again, maybe that might refresh Flash on your – on your computer.

Johannes Koch: OK.

Suzanne Theberge: All right. So, this measure has passed, 13 yes, zero no.

(Christopher Saigal): Great. That's a great review, by the way. Thank you for (reading) us through that.

Good. Is Andy on the line?

Andrew Baskin: Yes.

So, are we ready to start the next one, then?

Christopher Saigal: (Yes. Go).

Andrew Baskin: You got the OK? All right.

Do you hear me OK, everybody?

All right. So, the next – the next measure is number 2065, gastrointestinal hemorrhage mortality rate. This is the AHRQ measure that measures in-hospital mortality when the principal diagnosis is gastrointestinal hemorrhage. I believe there is also one case where if the initial diagnosis is primary is (inaudible) (varices) that the hemorrhage can be the secondary diagnosis. I believe that's a part of this measure as well. It's just a nuance to it.

Who is the presenter for this one?

Liliana Bordeianou: It's me, Liliana.

Andrew Baskin: Great. Liliana, you want to go ahead with this? Or ...

Liliana Bordeianou: Absolutely.

So, you summarized it very nicely. This is a hospital-level measure. And it is a new measure. And that is being proposed by the Agency for Healthcare Research and Quality. And, in terms – and they've done a very thorough testing of this measure, I would say.

The measure is based on administrative claims. And if you go to page 25 of the Web link, you will see the testing that they performed to look at the reliability of the measure.

Essentially, what they did is they looked at 458,307 patients using (prior) date. And they performed a reliability testing called signal-to-noise ratio which, I understand, is a very reliable way of describing how the in-hospital variation and (tier) might – I'm trying to explain this the best I can.

Of course, I'm not a statistician. But, I'm told that this is a very reliable way of testing for how clean the score, the ultimate score is. And if there is a statistician in the crowd, then you can, perhaps, go into more detail. And I'm told that you can get a score anywhere from zero to one. And if you end up, during this end, somewhere in the middle, then that's a good score.

And, if you go to page 28 of the Web site, you can see that they essentially saw a score that ranged anywhere from 0.01 to 0.6. And the average is 0.46. And that is considered a reasonable, reliable way of quantifying whether or not this is a clean measure or not.

And that is all they did for real – for reliability. The rest of the testing was all validity testing.

I'm done.

Andrew Baskin: Anybody have any questions about reliability testing?

If there are no questions, then we can go on and vote about reliability testing.

Liliana Bordeianou: I would also add that there seems to have been no technical review by the NQF, I presume by a statistician that who tells – who said in the notes to us that this particular measure – way of reliability testing is considered very appropriate and that they – the fact that they included, essentially, all hospitals ahead in their database is quite impressive and that this, therefore, should be considered at least moderate and maybe even high in terms of reliability.

(Christopher Saigal): Just (a clarification), when you say on the description of this, in-hospital variances, the cases in which a mortality (is) described to (inaudible) hemorrhage and it actually wasn't a (inaudible) hemorrhage, the cause of

mortality, that would be variance – the part of the variance? Do you know about that?

Liliana Bordeianou: No.

(Christopher Saigal): No. Any (inaudible)?

Andrew Baskin: Is there anyone of the phone from AHRQ that can help explain that to us because I think it gets into the statistical issues that probably challenge many of us.

Jeffrey Geppert: Sure. This is Jeffrey Geppert from AHRQ.

So, this measure of reliability, this signal-to-noise ratio, is essentially looking at how much of the total variation in performance that you see in this measure – how much of it is due to sort of systematic variation across hospitals and how much of it is due to, essentially, random variation attributable to the fact that some hospitals have very few cases that, for a condition like this, most hospitals, you know, don't have that many cases.

So, when you've got a relatively small number of patients in a hospital, you've got some sampling variability that you need to take into account. And that's, basically, what this metric for, to assess how much of the variation is due to sampling variability. And when you take that into account, then you have a more reliable measure in the sense that a measure that is less susceptible to that type of sampling variability.

(Christopher Saigal): But, it's not really looking at whether the actual – you know, if you looked at this measure from a different data source or look at how – (the fidelity to the) actual event in question, that's not been measured. This is a variation of the sampling strategy.

Jeffrey Geppert: That's correct. Yes.

Andrew Baskin: So, this is not – this is not reliability of data elements.

(Christopher Saigal): (I'm taking that's what that means). All right.

Jeffrey Geppert: (Inaudible) (that means it sounds like).

Liliana Bordeianou: So, I mean, we're measuring death. So, it's hard to imagine that the different data source will be less reliable. I think validity is different.

Jeffrey Geppert: The way to describe it is some sort of – some assessment of criteria in validity, you know, what's the (specificity) and sensitivity of the specification with respect to some external (gold) standard, not really reliability.

(Christopher Saigal): OK. (Inaudible). Thanks.

Andrew Baskin: Any questions or concerns before we vote on the reliability?

Well, the – I guess, we're voting – (commence the voting). And, so, we can continue.

Suzanne Theberge: Johannes, can you see the vote this time?

Johannes Koch: Yes. I did.

Suzanne Theberge: Great. OK. So, we have 12 votes.

Andrew Baskin: So, we have 12 voters or 13 for this? Thirteen voters for (inaudible).

Suzanne Theberge: Thirteen. We're all set.

Andrew Baskin: Great. The, I guess, we can move on to validity.

Liliana?

Suzanne Theberge: Sorry. I just need to read off the votes for the transcript after each vote. So, we have 13 moderate. And, now, we can move on.

Liliana Bordeianou: OK. So, validity testing. If you move on to page 28, you can start seeing the result of the validity testing. And (they'll want to) – it's easier to understand.

Essentially, what they did is something called empirical validity where they team up with several hypothesis as to what the relationship should be between

the score and what they see. And, then, they actually did the scoring and see whether that hypothesis (bore out) or not.

And one of the hypothesis was that high-volume hospital will have better outcomes. Another hypothesis was the patient – that the hospitals that transfer outpatient with GI bleeds all the time may have worse outcomes because they have less experience taking care of these patients and have higher mortality because of that.

And, then, on the next page, you can see a table and it is followed by the interpretation of the results. And, essentially, it appears that the hypothesis, when the testing was done based on (prior) data for the rates of mortality in various hospitals, there was indeed a negative association between the hospital risk-adjusted mortality and the hospital volume and opposite relationship between mortality and transfer outs, which fit with their initial hypothesis. And they used that to say that the measure must be valid because these hypothesis were (borne) out.

They also looked at the exclusions. And, essentially, there is only one exclusion that they are proposing, which is the exclusion of the patients that were transferred out to an acute hospital because, obviously, they are (not) dying within the hospital where the patient presented. And what – and what the analysis showed is that the hospitals that have transferred patients out actually have worse mortality than the patients that they are keeping.

In other words, they should keep sending more patients out because the ones they keep – that they kept seem to be doing poorly. And that was used – that's another point that they were trying to make within their validity analysis.

And, the, finally, when adjusting for outcomes, in addition to adjusting for comorbidities, a proposal was – this particular measure is trying to, also, adjust for the transfer in status. In other words, the patient that received the hospital – the patient (that transferred) from the hospital that just sent the patient away perhaps (in more extremist), perhaps sicker, et cetera – et cetera should be given credit for accepting (patients in the late stage in their care).

And they showed in the analysis later on – next page – that the – that this is justified because, indeed, patients that are being transferred in seem to do worse than the patients that arrive and take – and taken care of in the same hospital from the beginning to end.

So ...

Andrew Baskin: So, they're excluding transfers in and transfer out.

Liliana Bordeianou: So, they are excluding completely transfers out. And, then, when they are doing adjustments – and that's for – when they are stratifying patients for comorbidities et cetera, they are stratifying by transfer in. So, those are not excluded. But, they give an – they stratified to account for the fact that they were transferred in.

Christopher Saigal: So, you mean, like a – kind of a risk adjustment.

Liliana Bordeianou: It's a risk adjustment. Correct. And when they do in the end is they calculate this statistical risk (mortal) discrimination statistics, (assist) statistics. And they say that it's 0.8, which is pretty good discrimination between hospitals by deciles.

And they also look at their risk certification analysis using the data and they note, as expected, that the patients with esophageal (varices) do much worse than all the other cases. So, there seems to be both good stratification ability as well as discrimination ability between hospitals by deciles.

So, I would say that this is a pretty straightforward measure and that measuring death seems like a pretty valid thing.

Andrew Baskin: Any questions or comments?

Zahid Butt: Yes. This is Zahid. I have a question regarding the patients that were transferred out. Could it be that the patients did poorly, that were kept in-house because they were too sick to transfer out? And would the severity adjustment have adjusted for that?

Andrew Baskin: Well, that – well, that – AHRQ answer that one? (Is their office on), if they can answer it?

Jeffrey Geppert: Well, the severity adjustment would account for difference in severity in the patients that actually presented at that particular hospital. And, then, the intention of the – of the risk adjustment for transfers in is to account for any (sort of) clinical characteristics of those patients that are being transferred from one hospital to another that are otherwise accounted for by, you know, the diagnosis codes in the – in the comorbidities and measures.

Zahid Butt: Right. My question was different, not the ones that are transferred in but those that tend to – there was an implication made that those hospitals that were transferring people out had a higher mortality than could, otherwise, be explained. But, (my own question) was ...

Jeffrey Geppert: Well, that was – it's ...

Zahid Butt: ... could it be because the patients were too sick to transfer and, therefore, they ended up dying more in the hospital that they stayed in and whether there was a risk adjustment of those cases that could have, basically, sorted that out?

Jeffrey Geppert: So, the statement about hospitals that transferred out cases tend to have higher mortality – so, that transfer out rate is across the board. It's not just specific to the cases that are on the denominator, those – of those measure. So, the argument is, basically, that hospitals that transferred out large percentages of their patients – the construct is that they, you know, tend to have less capability, probably, and less quality of care.

And, so, the association between their overall transfer out rate and higher mortality rate for this particular measure is intended to be evidence of consistency with that – with that construct. But, it's not – it's not a measure of that patients who were transferred out who have GI hemorrhage who have a higher mortality rate. That (happens to be the construct).

Patrick, did you have a comment?

Patrick Romano: Yes. This is Director Romano – Patrick Romano from the AHRQ team. One other point is just the concern is often raised that hospitals may achieve low risk-adjusted mortality rates for conditions such as GI hemorrhage by transferring out all the patients who are going to die, leaving them with risk-adjusted mortality rates that look good. But, they essentially dumped their dying patients to other hospitals.

So, this is reassuring that, in fact, the hospital that tend, on average, to transfer out more patients still have higher risk-adjusted mortality rates with this measure.

Does that make sense?

Zahid Butt: Yes. So, that's sort of the other side of the argument. I was just simply sort of trying to sort out that patients that were being kept had a higher mortality rate and whether that risk-adjusted higher mortality rate would have actually adjusted for the fact that these patients were, perhaps, too sick to be transferred.

Patrick Romano: Well, it's possible that we don't know, of course, because the reason why these patients have to be excluded is that the program is designed so that hospitals and hospital systems can implement it using their own data without knowing what happened to the patient after transfer. So, essentially, we have a missing outcome problem when the patient is transferred.

We don't know on the individual patient basis what their outcome was at the second hospital. So, that kind of forces our hand to exclude those patients. But, fortunately, it doesn't seem to introduce bias in this case.

Zahid Butt: OK. Thank you.

(Edward Gill): I have a separate question. I guess it was alluded to a little bit earlier. And that this measurement tells us that the measure behaves in a way that we might have expected it to behave and there is some – and, obviously, that supports validity.

I guess what I'm concerned is, is there some body of evidence out there to – that would say that the diagnosis that was given is actually accurate and that the primary diagnosis really was a GI hemorrhage and was the primary cause of hospitalization?

Patrick Romano: Yes. I can address that, too.

So, I would say there's two bodies of evidence. One is that there are general (studies) that have looked at the validity of the principal diagnosis which is, of course, the diagnosis that principally drives the (DRG) assignment. It – therefore, it's a diagnosis that a lot of people are concerned about, including hospitals, CMS, other payers and so forth. So, in general, we know that the accuracy of the principal diagnosis is on the order of 95 percent, at least at the level at which we are using it here, which is an aggregation of principal diagnoses.

The second body of evidence comes from studies that have used administrative claims data linked with other data sources, their chart review or registry. The study that we cited in a – in a previous submission was a study that showed 88 percent (house predictive) value for GI hemorrhage diagnoses appearing in any diagnosis field on the hospital discharge abstract.

So, 88 percent (possible) value for all the diagnoses. And, again, we know that that tends to be higher. But, the principal diagnosis, which drives payment and which is the key diagnosis, is used to find the denominator for this indicator.

(Edward Gill): Thank you for that.

Andrew Baskin: Any other questions or discussions before we go to vote on validity?

I think we're ready to vote.

And I'll give everybody three seconds to change their vote. And, then, we have 13 votes.

OK. (It's like) we're locked in, then.

Suzanne Theberge: All right. That's 2 high and 11 moderate.

Andrew Baskin: And, then, we could move on to feasibility.

Liliana, you are doing great.

Liliana Bordeianou: Thank you.

So, feasibility. Because this is a collection of billing data, presumably, it will be pretty straightforward to gather the information from all the hospitals that are submitting their scores. And the clinical data about death and no death is hard to cheat on.

There is – I have some concerns about cheating on what you code as a diagnosis if you know that you are being measured for deaths from bleeding. And can you, then, say, “Maybe this patient died from a heart attack” instead or something else. In other words, cheat the system.

But, assuming that everybody is honest and that this is electronically gathered, it is still quite a feasible measure overall. It's what I would say. And it looks like our reviewers agreed with that. And the scores from everybody prior to this discussion on the high side and medium – moderate.

Andrew Baskin: Anybody have an additional comment to that? It was pretty straightforward.

I think we can move on to vote on feasibility.

I don't think we include (gamesmanship) as question (inaudible). But, I appreciate your concern. I have the same concern.

Liliana Bordeianou: Well, you know, as we start measuring one thing, people are going to try and (inaudible).

Andrew Baskin: Yes. The diagnosis – yes. The – which diagnosis coming more commonly seems to chance as we measure.

Liliana Bordeianou: Correct.

(Reva Winkler): Has everyone voted?

Andrew Baskin: (Well), we've got one more vote missing here. It looks like – anyone not vote yet?

Suzanne Theberge: Anyone have trouble voting?

Is anyone abstaining from this vote for some reason?

All right. It's possible.

Andrew Baskin: Somebody may have dropped off or walked away.

Suzanne Theberge: No. I – sorry. Was that someone?

All right. Well, we have 10 high and 2 moderate.

Andrew Baskin: OK. Then, we will – we will move on to usability.

Liliana Bordeianou: So, this is a new measure. So, it hasn't been used before for public reporting and – but, it does seem like a very credible measure that could be used in public reporting and could be used, if not (gamed), for improvement of (tier) or multiple (fronts), not just by (healthcare) or somebody (who wants GI suit) but how quickly medicines were administered, how quickly were they were (inaudible) of the emergency room.

So, it's a nice system measure of the quality of the (tier) within a hospital. So, I think it's both usable and could drive improvement. And I think that the benefits, even with a (gamesmanship) that potentially could occur outweighs negative consequences of (blaming) and naming some hospitals as providing lesser care.

Andrew Baskin: Comments? Discussions?

Patrick Romano: If I could make a slight correction? This is Dr. – this is Patrick Romano again.

Andrew Baskin: (Inaudible).

Patrick Romano: I think that this indicator, actually, has been in use, although it has not been NQF-endorsed. And it has been modified personally in response to the suggestions that this steering made in the – in the phase one review.

But, if you refer to page 14 and 15 of the submission, I think there are some examples given of previous uses and ongoing uses of this measure although, again, it has not been submitted previously for NQF endorsement.

Liliana Bordeianou: (Inaudible).

Andrew Baskin: Thank you. So, currently in use and with some significant historical data.

Zahid Butt: I have a question. This is Zahid again. Is the sort of recommendation to use this in sort of public accountability setting as only a severity-adjusted measure? Or, since it's such an important outcome measure – or, can it be used either as an adjusted measure or as an unadjusted measure? Because, I think, that makes a big difference, from my standpoint, in terms of its usability for public accountability.

Jeffrey Geppert: NQF endorsement is linked to the specific methodology that we have presented here. So, it would be linked to the risk adjustment and the reliability adjustment procedure as they have been presented.

Zahid Butt: OK. So, it will only be a risk-adjusted measure based on the methodology?

Jeffrey Geppert: Yes. That's the only one that's being endorsed. It's that this is the risk-adjusted methodology.

Zahid Butt: OK. Right. Because I know that it has been used not in public setting but in other settings as an unadjusted measure as well. So, I think that's good that it's risk adjustment.

Andrew Baskin: Thank you.

Any other comments or discussion?

And I think we can move on to usability (inaudible) going to vote – voting?

NQF's line dropped. We are dialing back (inaudible).

Reva Winkler: Here we are.

Suzanne Theberge: We're back. Sorry about that.

Andrew Baskin: That's quite all right. (Inaudible). We need to refresh the (auto vote, if you do that)? Or, I see this message (here).

(Edward Gill): Just the – (just the phone. We could still be in one the webinar).

Andrew Baskin: OK.

Reva Winkler: So, was there still an outstanding about ...

Andrew Baskin: No. We're ready for usability voting.

Reva Winkler: Great.

Andrew Baskin: There were no more discussion items.

Reva Winkler: Great.

Reva Winkler: I think we (did lose) ...

Suzanne Theberge: OK. I think we – I know some people had to sign off a little early. So, we may have lost somebody. So, we've got 11 high, 1 moderate, for a total of 12 votes.

Andrew Baskin: All right. We're done with that.

Well, Liliana, thank you very much, especially the statistical part of your (inaudible).

Liliana Bordeianou: Sorry.

Reva Winkler: Andy, would you have one more vote as the overall vote for endorsement?

Andrew Baskin: I apologize. But, thank you, anyway, Liliana.

So, vote for the endorsement – (if it meets) the criteria for endorsement.

It's like we have all the votes in.

Suzanne Theberge: Yes. All right. That's 12 yes, zero no.

Andrew Baskin: So, the – we'll move on to the next one, which is – is the next one (inaudible)?
Is that where we're going next?

Reva Winkler: Yes. (Right).

Andrew Baskin: I just want to make sure I got the right order. Let me just pull it out for a
second and see which one (is a two).

Johannes Koch: That's the normal colonoscopy.

Andrew Baskin: This is the normal one?

Johannes Koch: Correct.

Andrew Baskin: OK. That's Phil?

Johannes Koch: This one's Johannes. But, I think Phil is there, too, also.

Andrew Baskin: I'm sorry. I – the voices ...

Johannes Koch: That's OK.

Andrew Baskin: Hi. So, this is the – this is the measure 0658, which is a measure of screening
– well, not screening – well, colonoscopy – yes, screening colonoscopy for
which a normal result – in other words, there were no polyps removed or other
reasons for earlier follow up.

So, normal result in an average-risk patient with documentation and a record
of a recommendation of a 10-year follow-up examination. And I believe this
measure is reported out with some exceptions so that you could actually
measure separately the number of those that have medical exceptions or
patient exceptions but that those have – so that they can be reported out as
separate numerators.

Phil, are you on the phone? I think you're probably the presenter.

Johannes Koch: Well, I think Johannes is – this is Johannes. I think I am ...

Andrew Baskin: Yes. (Inaudible).

Johannes Koch: I think Phil – I think Phil is doing the next one, which is much more controversial.

Andrew Baskin: I apologize.

Johannes Koch: No. I got the soft ball. And I'm sorry (inaudible).

Andrew Baskin: All right. Johannes (inaudible). This is all yours. We've been hearing your voice. Great.

Johannes Koch: No. So – and this is either because this seems a lot simpler to me than all the other ones that we've already discussed. Or, maybe, I missed some of the pertinent parts. But, I did miss a little bit of the introduction.

So, to me this is a – you know, a renewal of an existing one. I think we've all voted and thought of this as a pretty transparent measure, which is that if you have a normal colonoscopy and no reasons for an exception, you shouldn't do another colonoscopy or shouldn't recommend another colonoscopy for 10 years.

And the variety of measure here, both on reliability, that we're going to start with seem to confirm that this is a useful and a reasonable way of addressing that and measuring that. So, while I am no – I am certainly not a statistician either. But, the numbers given on pages 29 to 31 on the reliability measures are – you know, for me, appear to be quite good on the testing that was done.

Andrew Baskin: So, let's look at the reliability testing first and the (inaudible).

Johannes Koch: That starts on page 29, at the bottom of it. And, then, it goes on 30. So, there is the number of events with the reliability as data set one and two.

Sorry. My phone line keeps dropping out, I think (inaudible).

Andrew Baskin: That's all right. We hear you.

Johannes Koch: OK.

Andrew Baskin: Yes. At least, according to the numbers here, the reliability looks, you know, extremely good. I don't know if anyone has a particular concern or question about this. If this – once again, it appears to be straightforward reliability testing.

No questions or concerns about reliability? So, I'm thinking that this is – now, you know, this whole thing about data element versus scoring, I guess I'm trying to understand what, technically, this is. Is this – I guess this is considered data element reliability?

Johannes Koch: So, I'm out of my league on that one. I don't know.

Andrew Baskin: Yes. And I'm probably stretching myself a little bit. So, does anybody want to jump in or is everybody comfortable with that?

Reva Winkler: Andy, the testing attachment should have that indicated by the developer on how they evaluated it. That's one of the question.

Johannes Koch: I'm sorry. I must have missed that. I don't know.

Andrew Baskin: Yes. I know – I'm actually in the document. I guess I went down a little too far. Reliability (inaudible) score is measured as the ratio of signal to noise.

(Pam Tourney): Hi. This is (Pam Tourney) with the AMA-PCPI. Can I – I just thought I could answer that question fairly quickly.

Andrew Baskin: Boy, I'd love you to go ahead and do that. Thank you.

(Pam Tourney): Great. It is at the score level and not at the data element level.

Andrew Baskin: OK.

(Pam Tourney): And that's indicated in (282.1).

Johannes Koch: OK.

Andrew Baskin: All right. Nevertheless, reliability looks very high.

Are there any questions or concerns about this testing or reliability in general?

Not to make it more complicated than it is, I think we can vote on reliability.

And I think we're supposed to have 12 of us still. Right? Or – we're supposed to have 12. We only have 10. Anyone not voted yet?

Suzanne Theberge: We have lost a couple of other folks.

Phil, I don't know if you have – I don't think you have Internet access. Do you want to vote verbally?

Nancy Faller: I am – it's Nancy. And I don't have the little boxes on my page anymore when you did that – whatever you did to bring the pages back up, my check boxes went away.

Suzanne Theberge: OK. What works the last time was just having you press F5 and – to refresh your browser. But, you can just vote.

Nancy Faller: (Inaudible) I have it back.

Suzanne Theberge: OK. Maybe, try quitting out of your browser and re-launching it. If you want to vote verbally for this one, hopefully, it will come up on the next vote.

Nancy Faller: OK. Moderate, please.

Suzanne Theberge: OK.

Andrew Baskin: Thank you.

Suzanne Theberge: And, Phil, did you want to vote verbally?

Andrew Baskin: I'm sorry.

Nancy Faller: I did. Sorry. I said moderate.

Suzanne Theberge: We have a moderate for Nancy. How about Phil?

Andrew Baskin: Is Phil on the line?

Suzanne Theberge: He was – he was there a minute ago.

All right. Well, if it – it seems like we've got enough votes. So we have four high and seven moderate.

Andrew Baskin: OK. Then, let's move on to validity.

Johannes Koch: All right. And, so, they are – similarly, the validity testing is quite high. The composite score is always over four with, you know, the overwhelming majority of people voting this as strongly agree.

Andrew Baskin: Now, this is systematic assessment of phase validity. I think we should note that. And similar to the methodology that we had seen earlier.

Any questions or concerns about this?

Reva Winkler: Andy, this is Reva. I just would like to committee's response to the comment that was submitted by American College of Physicians about the upper age limit in this measure.

Andrew Baskin: OK.

Reva Winkler: Or the lack thereof, actually.

Andrew Baskin: Yes. So, there is no age limit. I have to go back and I – you know, and I read the comments and then forgot to look again.

So, this is (inaudible) – OK. Yes. Well, if you want the committee to respond to that – or, I guess, the AMA already responded to that. Right?

Reva Winkler: Right. But, it – I think it's something you want to factor into your discussion of the validity, particularly as how the specifications are aligned with the evidence.

Andrew Baskin: All right. So, can someone who's (GI) remind me? But, I'm not mistaking, there is – the recommendation is that this 10-year colonoscopies would not occur after age 75. Is it?

Johannes Koch: So – yes. This is Johannes Koch. This – you know, it's a little bit of a moving target. Right? So, within the last year, the whole Choosing Wisely campaign has endorsed discontinuing colonoscopy in patients who had a normal exam by or before age 75. That, in and of itself, hasn't been incorporated into the guidelines and/or other work done by the three main GI societies.

So, you know, what – we don't know, for example, if a person has had an – it would – that's the – that's the – that's where it's moving towards. I think – you know, this – in my opinion, this is catching it right in evolution. Right?

So, when we talked about this last fall, that really wasn't in the forefront. It's certainly in the forefront right now. And I don't know how to – what to recommend – that it's likely going to be the standard. I don't know that it's been de facto accepted as that.

Andrew Baskin: So, I don't know how to deal with that because this is – essentially, we're not – (we're re-talking) evidence. I just wonder if one way to handle this would, say, that the score be stratified to those ...

Johannes Koch: That's (inaudible).

Andrew Baskin: ... age of 50 and 65 and, after 65, would be a separate account because there are many people who would believe that once you're 65, you don't need another colonoscopy at all.

Johannes Koch: Right.

And I've – and, then, the idea of the evidence reviewed to determine what do you do with people over 65, I think, could be handled by the user. I don't know if that makes much sense to people of if the AMA wants to comment since they are on the line here.

Zahid Butt: But, I think – this is Zahid. I think Johannes is correct that, you know, once it sort of becomes a de facto standard, if you will, perhaps in the next go around, it could be somehow incorporated. But, currently, really, that's sort of the guideline in terms of who over the age of 50, which is kind of what this is really primarily addressing.

Andrew Baskin: Will Zahid or Johannes or anyone who is more familiar with the GI – I don't recall the evidence review. But, in the evidence that was cited, did the evidence stop at any particular age or the evidence was, basically, no upper age?

Zahid Butt: I don't believe that the guidelines have that upper age in there at this time with the evidence that was cited.

Johannes Koch: It don't.

(Richard Luetkemeyer): I think there is a caveat that if you're over the age of 75, the life expectancy has to be equal to or greater than 10 years.

Andrew Baskin: Yes. I guess the concern is that when we get to the idea (of quickly be doing) harm to that portion of the discussion, there are some that may be concerned that even a healthy, you know, 75-year-old, you could potentially be doing more harm than good by doing a colonoscopy. And nobody knows the answer to that. But, it would start – (if the patient gets over), that would start to be a concern of mine.

So, I don't know how we'll address that. I guess we can address it at that time. I guess we have no definite answer except to say we'll base our decision on the measure as presented. It's the current evidence that was used to develop this measure and we accepted that evidence.

Johannes Koch: Yes.

Andrew Baskin: If there's no – if there's any other discussion, then, please. But, if there's not, then I think we should be voting on the – on this validity based on the measure as is. And, during the usability, if we want to discuss potential and untoward effect, we certainly can bring it up then.

Johannes Koch: OK.

Andrew Baskin: No other comments, then?

OK. Then, I think we should go ahead and – with the vote on validity.

Suzanne Theberge: Nancy, are you able to vote this time?

Andrew Baskin: We have 10. How many did we have (referenced)? We have 11?

Reva Winkler: We had 12 last time. We had 11 last time. You are right.

Andrew Baskin: Even with Nancy, we have 11. So, I think we have the full amount.

Suzanne Theberge: All right. Yes. So, that's 10 moderate, 1 low.

Andrew Baskin: OK.

Then, I guess the next is feasibility.

Johannes Koch: All right. And there – again, I believe that, you know, the metrics used show that this is a – this is a metric that is measurable and incorporated in the majority of reports and is very feasible.

Andrew Baskin: When we talk about feasibility – so, this is something that, I presume – I don't think there's any coding for this. Is there? I don't think there is a CPT-2 code for this. So, this is – this is, essentially, chart review?

Johannes Koch: Well this, I think, is – much of this is based on the data elements that were defined for the AGA and the GI (quick) outcomes registry. So, some of that discussion, I think, will sort of take place more so in the eMeasures discussion which, I think, is going to follow.

Andrew Baskin: All right. OK.

Johannes Koch: But, to the extent that this is currently, you know, mostly reported through this registry mechanism, it's feasible to the extent that people are reporting this data and, you know, getting measure performance out of it.

Andrew Baskin: OK.

Reva Winkler: Yes. Andy and Zahid – it's Reva – in terms of any discussion around the eMeasures, I mean, it should be even incorporated into this whole discussion because it was presented as part of the same measure (division).

But, I also want to just point out time, Andy, and how you might want to handle the fact that we are very close to our time and we need to, at least, do some public comment.

Andrew Baskin: So, do we want us – I think eMeasures may take a little time since we've not talked eMeasures and I know Zahid has done some expert analysis for us. Then, I think, we should probably halt the voting at this point, being that there is only three minutes left and there'd be public comment. And, when we return, we'll go back to feasibility. But, we may have to back up to the eMeasures specs and make sure we have that straightened and that our prior voting takes that into consideration.

Suzanne Theberge: Yes. We can pull that up to start with on the next call next Monday. And if there is no other questions or comments from the committee at this time, then we'll go to public comment.

Does anyone have any questions or comments?

Andrew Baskin: Go ahead and open it up so the public has an opportunity.

Suzanne Theberge: (Natalie), can you open the lines?

Operator: All lines are open.

Suzanne Theberge: All right. It sounds like ...

Andrew Baskin: I was going to say – so, it sounds like we're going to give everybody a minute and a half back because I don't think opening up another question's worth – we don't have the time for that.

Suzanne Theberge: All right.

Andrew Baskin: I think things went pretty smoothly. So, thank you to everyone on the committee and in NQF. This was more systematic than I had ever hoped for. So, good stuff. Now, we know what to expect next time.

Johannes Koch: And I – this is Johannes. Just one least comment. I'm sorry. I have to apologize. I have to go away for – and I'm going to miss the next call because I have a family issue that I have to attend to next week. So, I apologize for that. But, thank you for allowing me to participate and look forward to the next time.

Andrew Baskin: Thank you. We'll fill in for you next time.

Johannes Koch: Thank you.

Suzanne Theberge: All right, everybody. So, we'll talk to you next Monday, April 8, again, at 3 p.m. Eastern Time, and we'll cover the remaining issues for this measure and, then, discuss the last three measures at that time.

And, if you have any comments before then, please don't hesitate to call or e-mail the project team and we'll look forward to speaking with you next week. Thanks you very much for your time.

Andrew Baskin: Thank you.

(Johannes Koch): Thank you.

Operator: Ladies and gentlemen, this does conclude our conference call. Please, you may now disconnect – you may now disconnect.

Suzanne Theberge: Thank you.

END