NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
March 20, 2014
12:00 p.m. ET

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

Reva Winkler: Hello, everyone. Welcome to the workgroup call for the cardiovascular project. This is Reva Winkler along with Lindsey Tighe and Vy Luong from NQF. We just want to check and see who of our workgroup members have joined us. I've heard from Joseph Cleveland, Elizabeth Delong. Liz, are you with us?

Elizabeth Delong: I'm here.

Reva Winkler: Great. Mary George is here. Kristi Mitchell, you're here?

Kristi Mitchell: Yes.

Reva Winkler: Great. George Philippides?

(Off-mike)

Reva Winkler: Not yet. OK, Henry Ting?

(Off-mike)

Reva Winkler: All right. Hopefully, George and Henry will be able to join us. I know that we have other folks from various measure developers and we'll be looking forward to hearing from you as we go through looking at the measures.

Thanks to our workgroup members who have submitted your preliminary comments to us. We've put them all into the worksheets on the individual measures for your review.

Today, our purpose is to have some preliminary discussion of the six measures assigned to this workgroup. In our two-hour call, we need to get some discussion on all six measures, so that puts us a little less than 20 minutes per measure. So, I'm going to ask the workgroup members to, you know, focusing in on areas that are potentially issues, concerns, problems for things that seemed very straightforward that everyone is in agreement, we don't need to spend a lot of discussion time.

So, with that, we have three measures to start off with that address the topic of AMI and the second three measures that are all about mortality for PCI. So, we'll get started with the first measure, number 0286, Aspirin at Revival. Our primary discussant is Dr. Mary George. Mary, did you want to introduce the measure and begin discussing your thoughts on the evidence?

Mary George: Right. So, first off, it's important to understand that this measure is an outpatient measure. And it is specifically the percentage of ED AMI patients or chest pain, probable cardiac chest pain patients, who received aspirin within 24 hours before ED revival or prior to transfer. So, this is not for those patients that are admitted. So, that's probably the main key thing to understand about the measure. So, citing the evidence for aspirin for AMI or probable cardiac chest pain, they use 2012, 2013 ACCH guidelines for UA, NSTEMI and STEMI.

Class one level I recommendations with – I really didn't see any issues with that evidence. They describe the relationship between aspirin at revival and reduced adverse outcomes. The question of the gap in here I think was one area that I think people should perhaps focus on a little bit. The 25th percentile for this measure is still at 100 percent adherence. And the 10th percentile, the 87 percent, national average 96.4 percent, so extremely high adherence to this measure with little gap in care.

They did look at it and stratified by race adherence for White, 97 percent, Black, 96 percent, Hispanic, 95 percent, so little evidence for racial disparities. It's, you know, obviously AMI and cardiac chest pain is a highly prevalent condition and costly condition, so I'll stop there, (it would be) significant if anyone has any questions.

Joe Cleveland: Mary, this is Joe Cleveland here and I think you've highlighted – I agree with you completely that, to me, as I read through this measure and I guess this has been – this is a measure maintenance practice not a new measure …

Mary George: Correct.

Joe Cleveland: … but I think, to me, the real significance is the gap in care, is there a gap. And that's where, I guess, I would – obviously, I think discussion is warranted in terms of with 96.4 percent. I mean, I guess the tradeoffs are if we say, you know, we've kind of topped out and this is no longer a measure to endorse because we're kind of, you know, there's just not a gap, you know, I guess, the downside is that people kind of take to write-off this and could there then be a backslide that I think the real critical thing is, as you brought up, which is the, you know, at the 25th percentile, it was already 100 percent adherence. So, that, to me, seems like the critical area discussion for this measure.

Mary George: Yes. Go ahead.

Elizabeth Delong: Liz Delong. I'd like to add that I think it's important to wave the tradeoff if there were the same gap in care but it was relatively easy measure, that'll be one thing. But the coding instructions for this measure are very, very complex and it looks as though – I got confused but it looked as though you have to hand-abstract the data which, to my mind, it's not worth the benefit.

Reva Winkler: Right. This is Reva. Just to sort a couple of things out, we certainly have seen a lot of these measures that have been place for a while become topped out. I just want to make the committee – the workgroup aware that NQF does have a status called reserve status that can be applied to a measure that seems to be topped out with little opportunity for further improvement as long as it meets all the other criteria and scientific acceptability, usability, and

feasibility. So that is one option if this is a concern and it's something the entire committee can certainly decide to do on your final recommendations.

Joe Cleveland: I guess, following up on Liz's point, again, as the feasibility with the amount of administrative coding that's required, and I agree with her. I think it seems like a lot of, you know, abstracting, so that'll work. So …

Reva Winkler: All right.

Elizabeth Delong: And not just the abstracting. For somebody to actually record it correctly so that it can abstracted, I actually wondered if people know all those rules and they're able to follow them.

Reva Winkler: Shall we ask the measure developer for any (on the phone)?

Dale Bratzler: So, this is Dale Bratzler. I'm here.

Reva Winkler: Hi, Dale.

Joe Cleveland: Hey, Dale.

Dale Bratzler: Hi, Reva. So, I actually think the abstraction is not a huge issue. So, this is just one measure out of our long list of outpatient measures and hospital outpatient departments are required to capture for CMS as a part of the hospital Outpatient Quality Reporting Program as defined in statute, and the hospitals have the mechanisms in place and have for many – for a number of years now to actually abstract and capture this data.

Now, I think we all would love to see most performance measurement to go to an electronic platform. And, in fact, there is a team working with CMS on re-specifying many of these measures for electronic medical records to eliminate the need to do any manual chart abstraction. But that is currently the system in place and there are more than probably 4,000 hospitals that actually capture this data now. So, that actually happens routinely and has now for several years.

Now, I can't argue too strongly about the whole issue of the measure being topped or, you know, potentially topped out. There are more than 4,000

hospitals. They captured this data, so the 10th percentile may represent a group of 400 or so hospitals that have performance rates that are fairly low.

This measure was actually developed originally as a part of a hospital transfer measure list looking – focusing particularly on patients who have chest pain that show up at hospitals and then subsequently get transferred to other facilities because there is already a performance measure in the inpatient status that if a patient comes in, let's say, with a heart attack, there is already an inpatient performance measure on use of aspirin for those patients.

So, this measure only focuses on those patients who come in to a hospital and are not admitted to the hospital and most of these patients wanted to get transferred to – from a rural hospital to a larger facility.

Reva Winkler: Mary, did you want to maybe touch on the reliability and validity testing of measure?

Mary George: Right. So, many of the things I was going to say, Dale actually went over. So, thank you, Dale.

There are 11 data elements included and they did data testing on these data elements. They said that the – they could use administrative claims, paper medical records, or electronic records for testing. Data testing is in 2012 data at the facility level. And I think the – probably the one concern and they – because they did the data element testing is that the data element level, it serves as the validity testing.

The developer made one comment that there maybe some difficulties in properly including or excluding patients with probable cardiac chest pain. Their – when they did the data element testing, at least on the data element aspiring received, 98 percent agreement with the CAP of 0.86. And other data elements that they tested have fairly high percent agreement as well. And they did testing on over 212 patients from over 200 hospitals, as I understood that correctly.

This measure does exclude patients that are on anticoagulant and it excludes patients with the documented reason for no aspirin on arrival. So, I think …

Reva Winkler: OK. Thoughts from any of the other workgroup members?

George Philippides: George Philippides. You guys here me?

Reva Winkler: Hi, George. How are you?

Male: Yes.

Reva Winkler: OK. So, Mary, anything about feasibility or usability on this measure?

Mary George: So, feasibility, they stated that it was undergoing electronic retooling which Dale Bratzler confirmed and noted the difficulty in capturing probable cardiac chest pain from the electronic fields. It's – currently, they used ED records and EMS records for the documentation. I don't know how often they would be using the EMS record, so they're oftentimes not incorporated into the medical record or lost, but they should have this information from the ED records. So, that was what I had on feasibility. I didn't seem to see any real concerns beyond the possible electronic coding that they mentioned for probable cardiac chest pain. Maybe the developer could comment on that?

Reva Winkler: Dale, are you there?

Dale Bratzler: So, I'm sorry, what was the question about the probable cardiac chest pain?

Mary George: There was a comment in the documentation that it may be difficult to capture that in electronic record.

Dale Bratzler: Yes. I'm not a part of – I haven't been a routine part of that workgroup. But, you know, I think that's been the challenge with a lot of the transformation of the current process of care measures to an electronic format, and that's much of that type of information often gets documented in text fields rather than structured field to the electronic medical record. So, I'm not surprised that they may be having some challenges identifying your appropriate fields in the EMR to identify that population of patients, the denominator population that has probable cardiac chest pain.

I think one of the jobs or somebody else from the team maybe on the call and they may have more knowledge of those ongoing discussions than I do.

Female: Yes, I'm on the call. You basically say that it's difficult because of the nuances in the paper based data element to transfer that over into a (store) based measure.

Mary George: It's possible that could be something that could be handled through education with coders and being able to pick that up – if it's picked through ICD codes, I would imagine.

Elizabeth Delong: And that's where I'm coming from, just the gap in care weren't the amount of effort that's going to be to collect the measure.

Reva Winkler: But that will be a primary discussion point (straightforward) to the in-person meeting. Anything else we want to talk about with this measure before we move on to another, Mary?

Mary George: Just in terms of usability, the measure developer correctly stated it's been publicly recorded for many years. It's part of the CMS Hospital Outpatient Quality Recording. It is for public reporting and payment as well as joint commission accreditation. And as really, my overall question looking at this measure was really with taking the time to abstract this measure, does that detract from other measures that might have greater performance gaps and deserves more attention.

Reva Winkler: All right. Anything from anybody else before we move on to another measure?

OK, hearing none, the next measure on our agenda is measure a 2377, Defect Free Care for AMI. This is a new measure. It is a composite measure. And Elizabeth Delong, I think, you're the primary for this one. Describe the measure and then talk about the evidence.

Elizabeth Delong: Sure. Let me introduce my opinions by saying this – I found this to be a very difficult and confusing measure to get through, possibly do the two reasons. One of which is that a lot of the text was repetitious. It the same – there was a

cut and paste mechanism, I guess, that answered several questions the same way as a matter of fact, and I don't have it right in front of me. There was something about aspirin – I'll find that soon.

But I think possibly the demand for filling out the measure criteria have overwhelmed the developers to the extent that they resort to cut and paste and answer questions that weren't asked in the same text that answers – that tries to answer the question. They also repeat the guidelines several times. It seems those could be consolidated into one set of guidelines with reference numbers like we do on manuscript.

That said, this is an AMI mortality measure and it has 11 components. I was a little concerned that the – most of the justification for this measure was as a whole and there was very little attention to the individual components. So, with – it's an all or none measure, in other words, all 11 components, all 11 opportunities have to be met. Actually, it was eight measures for STEMI and 11 – eight for non-STEM, I think, and 11 for STEMI, but it depended on the individual patient and whether they – that opportunity actually pertained to them. But that was an all or none so that each patient gets a zero – each patient is given a zero or one depending on whether all of their measures were – their qualifying measures were fulfilled. So, for example, counseling on smoking is up there with aspirin at arrival.

So, that's my main concern about this measure in addition to the fact that somehow the form was filled out in such a way that I was reading things in one section that didn't pertain to that section.

In terms of a gap, there does seem to be a gap, but once again, that was over the whole measure and not over the individual component. So, it's hard to say which components actually – which should be weighed more than the others and whether we really need all 11.

Reva Winkler: All right. Do we have anybody from the measure developer on the line?

(Tracy Conley): This is (Tracy Conley) from the ACC. Dr. Masoudi was trying to get on. He wasn't able to get through.

Reva Winkler: Really?

(Tracy Conley): Yes.

Reva Winkler: (You mean), to try?

(Tracy Conley): Yes, he was trying to get in and hopefully he's redialing back in.

Reva Winkler: OK.

(Tracy Conley): But I will take your comments back to him and we can prepare for the face to face. I think these – there are 11 measures based on the STEMI and NSTEMI. They are not weighted, so they are all equal whether the patient got care, got the defect free care, that is correct. So, no – a lot of attention probably was not paid each individual measures and it was committed as a whole composite measure.

Henry Ting: So, this is Henry Ting. I would just (inaudible) to Liz Long. And I had a few comments to add. I agree with Liz's comments so far. And a couple of comments that I would add in addition to Liz's comments is, one, aspirin prescribed at discharge. It (say) to maintain the dose of 75 to 152. I don't think we have 75 milligrams in the US. I think the smallest dose we have is 81. So, that's one comment to consider.

(Tracy Conley): OK.

Henry Ting: The, you know, I had a question also about sort of, you know, so (inaudible) would have to join ERG action (inaudible) right now. Is that …

(Tracy Conley): Yes. So, that's actually been – being reported currently in action and has been probably for the existence of the registry.

Henry Ting: OK. So, my question, I don't know if it's a concern or a question, but it's – you know, currently, there are 5,000 hospitals in the United States who sort of treat AMI patients, and only about 1,000 hospitals, give or take, a few hundreds here or there are actually participating in action. So, how does – is this a measure for accountability or just for quality improvement? How is this

measure going to be rolled out and used if only 20 percent of hospitals are currently participating in action?

Another question I have was, you know, sort of excluding hospitals with less than 25 (portal) tax a year. You know, I don't know if it's in the rationale because, you know, you say you're worried about low volumes, but, you know, from a quality perspective, it's probably specifically at those low volume hospitals that we probably have the most gaps in quality of care or processes of care, particularly the ones we're looking at. The low volume hospitals don't see this often. And to exclude them, you know, we just want to be sure that is the right thing to do.

And I think those are my major comments for the group to consider.

(Tracy Conley): OK. Dr. Masoudi, were you able to get on?

Frederick Masoudi: Yes, I'm here.

(Tracy Conley): OK. Were you – I don't know how much you heard or …

Frederick Masoudi: Yes. I mean, I think, you know, ultimately, the purpose of getting NQF endorsement for any measure is the possible participation in the future in the purposes – for the purposes of public reporting. Although, again, you know, I think the reporting strategies have yet to be determined. I mean, I think that, you know, again, with voluntary public reporting processes, having 1,000 hospitals and catching participants would be relatively – could provide relatively robust recording programs for participants within that registry.

As far as the dosing of aspirin is concerned, that's just concordant with CMS specifications of measure sets, has been issue of concordance, so I think it's a relatively minor one in that, any of the reasonably used doses of aspirin will qualify for success. That is to say, by broading the (inaudible) but I don't think it does any harm.

And I can't remember what the third comment was.

Henry Ting: Fred, this is Henry. Excluding hospitals of less than 25 MIs from the measure.

Frederick Masoudi: Yes. So, you know, in part, that's an issue of ensuring that we have, you know, robust sample size for measurements and we can certainly take that under advisement.

Henry Ting: The concern probably maybe is that those hospitals are low volume that would have the biggest gaps in these processes of care and making sure they're providing them.

So, I mean, Fred, this is really maybe not a question for the ACC but a question of if NQF endorses this and that all 5,000 hospitals need to join in action in order to, you know, but most of the measures were (consuming us). We've been told that's for accountability and payment.

Frederick Masoudi: Yes. Well, I mean, I think that, you know, NQF endorsement is certainly the first step for accountability. However, there's no mechanism whereby anyone would be forced to participate in the registry for purposes of the public reporting program unless that were, you know, unless it were widely, you know, or legislated or mandated to be widely adopted. So, I – so, I'm fairly sure I understand the concern.

Henry Ting: The concern being is, you know, in our work (lab) in this room so far, we've been told that we're looking at measures that are intended for accountability which is payment. So …

Frederick Masoudi: … accountability – actually, accountability could also be involved a voluntary hospital level public reporting program such as that run by the STS and that initiated by the ACC. And voluntary public reporting program has nothing to do with payment.

So, there are a variety of different levels whereby accountability, you know, where things would be considered accountability. There wouldn't be necessarily payments. And those payment issues …

Reva Winkler: Yes, this is Reva Winkler. I just – to echo Dr. Masoudi's comments, when we're talking about accountability, we aren't specifically indicating payment. That's one possibility. However, accountability could include public (listing), accreditation, credentialing, you know, and some of these other accountability applications. So, payment is not the only one.

Henry Ting: And just to be clear, as we endorse this measure, it is for accountability and it is publicly reported, so not linked to payment. Then hospitals who haven't joined ERG would be sort of – have a public reporting of zero percentage or no data or how would they be recorded?

Frederick Masoudi: Well, I think that, you know, in fairness, Dr. Ting, I don't think any of the reporting structures have been completely worked out for many of these measures. So, I think, you know, the application doesn't request specific issues around reporting structures and how those things will be done. You know, again, the STS and ACC have public – have voluntary public reporting programs whereby participants in the registry can opt in to reporting. And if they don't opt in, they don't get reporting. And if they do opt in, they end up having a report. And that's how that – that's how that particular program works.

But again, and I don't know that the details of how a measure is going to be implemented in the purpose of accountability is part of – is at least, in my understanding as part of the application, isn't much like part of that. But I could be wrong. I'd of course defer to the NQF on that.

Reva Winkler: Yes, this is Reva. I mean, it's almost impossible for committees to avoid thinking about potential applications. But again, we can't foresee the future in terms of specific program. So, really, we're asking your evaluation on suitability for use in any of these potential programs.

Elizabeth Delong: Fred, this is Liz Delong. I had a couple other comments. One of which, and it's sort of repetitious to what I said before, the individual components unevaluated in terms of reliability or in terms of their performance gap. Also, I didn't see where you would address missing values. And I would think the burden of missing values would be variable over the components.

Frederick Masoudi: Yes. I mean, as it turns out, I can't – you know, I have to look back at the individual components. I think some of those had to do with the way the applications are configured. And again, staff from NQF can correct me if I'm wrong. Each of these, you know, many of these individual metrics have at some point in the past been evaluated on these very issues, but that hasn't been included in the application program because – the application materials, again, because there are 20 different component parts. Of course, we'd be willing to, you know, we're willing to share things as required by the panel.

In general, missing this on these particular data elements specifically those in term of whether or not a medication was prescribed or whether or not a contraindication existed are very, very small for each – for each specific process of care. And that's in part because these registries are designed to assess for just that. But, you know, from empiric level, the level of missingness in terms whether a process of care was administered or not or whether or not there was a contraindication and it's quite complete, typically well above 95 percent.

Elizabeth Delong: I would still worry about the variability so that it's possible that some of the more important components would be missing at a greater frequency and the score might be a little biased in that respect.

Frederick Masoudi: Can you – can you be a little bit more concrete about that perhaps?

Elizabeth Delong: So, if you're – supposed beta blocker discharges presumably the most important of the component, but that's missing a fair amount of time. It has – or a relatively large amount of time given the spread in percent of missing values. Then patients for whom that is not recorded are not assessed on that particular one that is possibly more important that others and the score, if your site – if a particular site basically leaves that one out, their score might be higher.

Frederick Masoudi: Right. I mean, I could be wrong, and the staff can correct me on this, but generally, you know, missing values for specific process of care are not generally excluded from the dominator but they're –

but if there isn't actual documentation of having complied with the process, that is considered non-adherence with that particular (inaudible). So, in that respect, I think that – and again, you know, correct me if I'm wrong, but that would – that essentially – I mean, if that's what your concerned about in terms of the missingness, I think that gets around that issue.

Elizabeth Delong: Yes, I missed that. Thanks.

Henry Ting: So, one of the question I had, Fred, and also for NQF is, if we look at this composite measure, does that mean the individual measures for some the same elements would still get endorsement or we could consider not including them? For example, the measure we just looked at with aspirin at arrival, and that's included in this composite.

Frederick Masoudi: Right. I can just speak to what happened in the past. Well I just defer to NQF staff on that issue. That's more a policy issue …

Reva Winkler: This is Reva.

Frederick Masoudi: … I could talk about the history of something (inaudible).

Reva Winkler: Yes, this is Reva. What we're seeing is a bit of an evolution. As we started out with the individual process measures, we've had success with improving performance, and there is a growing interest and demand for composite measures that combine multiple elements. And so, we are still – NQF is, you know, accepted both types of measures. So, the question is the utility of one versus the other. I think most of the components, there is a – some version of the measure endorsed per hospital caring for patient with AMI. But this is collected through a different mechanism. It's constructed as an all or none. So, it's not absolute one on one connection. So you want to want to evaluate this measure on its own merit in terms of what it actually measures, any information that it provides.

As we – as this committee goes through all of the other individual measures, you will do the same for them as well.

Frederick Masoudi: And so just to clarify that if history exists, in example, Dr. Ting, with the CathPCI registry, there were – and this was discussed on the last call, there's a composite medication metric that was up for re-endorsement. And there was a question as to whether or not – as to why we hadn't applied for individual measure endorsement, and that was because during this previous cycle, we were asked to bring those into a composite and did not get individual approval for the component measures but were given approval for the composite.

Henry Ting: Yes, and I like that. That's terrific, Dr. Masoudi. It's – it seems like it would – to look at this individually may not be sort of the best thing on its merit. If you're going to have 11 individual measures, for example, plus the composite, getting all that endorsed, I just don't think that would be useful in use of anybody's time to collect at the hospital level.

Reva Winkler: Were there any other comments on this measure? I think we've touched on the evidence. Liz or Henry, was there anymore on evidence because I think you also talk about reliability and validity, certainly the usability and feasibility, was there anything else you wanted to raise before we move on to another measure?

Elizabeth Delong: My only other note is that I couldn't find how exclusions were determined. But that maybe that there was so much text there, I probably miss something.

Reva Winkler: Fred, did you want to address that?

Frederick Masoudi: Yes. I mean, exclusions are generally determined by the documentation of contraindication to the individual component measure.

Reva Winkler: All right, anything else?

All right, then, next on our agenda is measure 2473. This is Hospital 30-Day Risk-Standardized Acute Myocardial Infarction Mortality eMeasure. And so, just to give you a little context within the NQF portfolio, I think everybody is aware that there has been a claims based 30-day mortality measure for AMI for quite some time. And there is, you know, movement as we talked earlier towards development of eMeasures.

And so, this is one of the first outcome eMeasures that we've seen. We – as part of the evaluation from staff, we did have it looked at by our folks in the RHIT group input. But this is a bit of frontier built, you know, adventuring for us with one of first outcome eMeasures.

So, Kristi Mitchell, I think you're the lead discussant for this. Did you want to begin introducing the measure?

Kristi Mitchell: Sure, thank you.

First and foremost, I think it's a fantastic development to take a measure that has been applied and used in public programs for a while and really transform it into something recognizing the ongoing investment and resources to collect data, you know, either manually or through registries, claims. If we can use electronic data, this is, you know, almost revolutionary.

And so, with that, in mind, this again is this Hospital 30-Day Risk-Standardize AMI Mortality eMeasure. And, you know, I think unlike the others, the whole importance to measure and report concept is – has been well established and that we are really focusing on the reliability and validity, I would think, of the – from an eMeasure perspective. And I believe if you had an opportunity to look more closely at this, the specifications and the whole eMeasure technical review concept is where really the action is and where all the questions, at least, the majority of the questions that I had reside.

And so, just to kind of fast forward there is eMeasure technical review, that's a requirement of this particular measure. There were specifications that included codes again to identify MI discharges, date of birth. The measure developer took the opportunity to translate the ICD-9 codes to ICD-10 codes and provided that in the specifications. These outcome measures in fact risk adjusted using the statistical risk model and the developer provided descriptions and values for the five variables used in the model.

(Off-mike)

Kristi Mitchell: Sorry.

And so, when we're looking at the eMeasure feasibility assessment, the developer provided an IT survey or – and quality expert survey. And my understanding is I think they interviewed seven responded. And I'm not certain what the standard or the bar or the requirements are from NQF. And I'd be curious, I'm actually going to ask that question. Is there any precedent for how many people should or could be surveyed or was this something that the developer did on their own?

Reva Winkler: And this is Reva. Actually, the NQF did do a project around eMeasure feasibility assessment last year and this was certainly a point of debate on how many. And so, really, NQF doesn't have an absolute number. I can tell you that the number that were used by this developer were really quite significant and many more than (what's sent to us) by that committee. So, NQF doesn't have any issues with the number of assessment that were made with vendors and users.

Kristi Mitchell: OK. I still think that there's some question in and around, you know, the type of question that was asked. The number of EMR systems and sort of the test and assumption that even if an EMR vendor, so that they extract the data from their system, as to whether or not the data coming out can be abrogated, if you will. Just created (that for) benchmark. So, I still have some ongoing questions about that in general.

Reva Winkler: Did you want to ask the measure developer to verify anything specific?

Kristi Mitchell: Yes, actually. Are you at liberty to discuss the five EMR systems that were interviewed or surveyed?

Reva Winkler: Do we have anybody from the measure developer on the line?

Susannah Bernheim: Yes. Hi. This is Susannah Bernheim and many of our team members are on the line as well. You're asking, can we say, who we in the …

Kristi Mitchell: Yes.

Susannah Bernheim: Is that the question?

Kristi Mitchell: Yes, who within this EMR system or which EMR vendor, and then the type of the – the type of person who did the assessment like who participated …

Susannah Bernheim: Yes, and so there was a couple of different assessment. I think your referring interviewing serving people who were on the IT side within hospitals, is that right?

Kristi Mitchell: Yes.

Susannah Bernheim: And we also did some, you know, direct assessment of the data that came out of the backend of the EHR. So, I don't have that handy. We actually did that in collaboration with another CMS contractor. We can – if it's something that's important for the committee to know, we can track back and see if we can bring that to the steering committee later.

Kristi Mitchell: I was just curious because, again, the number is so small of the folks that you're interviewing. I just want to have a better sense of this. Within those five EHR systems like what is the, you know, like CIO – like who was the person responding to this?

All right. And that could be (labored out). The technical review, as they said, was found to follow the industry accepted format, you know, for eMeasures. The measure logic that was captured was automated and it seemed to comport to the measure calculation. There's been – sorry, I'm jumping through my notes here.

Looking at the reliability testing, I think this is the one area that I had some question about the validity of using action get with guidelines – excuse me – action registry get with the guidelines as a mechanism for testing and retesting those data elements that were to be extracted from an EMR and whether or not we think that there's enough similarity between those data elements to make that or an appropriate mechanism or method. So, I'm asking the developer what was the thinking and using the action get with guideline to registry data.

Susannah Bernheim: Right. So, it's a good question because this is an unusual approach to measure development and have to do with service. As you can imaging, the state of access to large aggregated EHR data at this point. So, our goal was to create a measure obviously that was intended for EHR use. We needed there for access to clinical, you know, key clinical data elements that could be used for risk adjustment and that could be feasibly extracted from an EHR. And so, we used action data because we knew that it had a large number of hospitals and auditing system and variables that were thought to be clinically important.

And the first step in the measure development process or nearly the first step was to look at all of the available data elements within the registry, and with a group of experts categorize each of them as whether or not they would or would not be likely to be able to be feasibly extracted from an EHR, and we have some very specific criteria there.

So, we essentially try to subset the action registry into data elements that we would use the risk model development that were likely to be successful in EHR testing. And then, as you know, we later went back to ensure that those variables indeed were not just through survey but also through running the final version of the – what they call machine readable version of our measure through the backend of EHR and comparing what our nurse abstractor would pull looking for those same data elements from the front end to confirm that indeed we're getting the variable we're looking for.

Kristi Mitchell: OK.

Susannah Bernheim: Does that answer your question?

Kristi Mitchell: Yes, it did. This is a question actually to NQF as well about sort of the standard for performance score validity testing, and whether or not sort of the preference of the sensitivity, specificity set versus what the developer did. Can you speak to me a little bit about sort of the preference of that particular standard over the other that was used?

Reva Winkler: Sure. It means – I think that if you look at the algorithm on scientific acceptability, while testing reliability validity is the level of the data element

which was done here, it's – looking at the results for the measure score, which of course is really the thing that's most of interest, would be preferable. And so that would allow you to rate it higher. But the evaluating reliability validity of the level of the data element is still perfectly acceptable.

And so, we do not specify to the developer how they do their testing. We let them, you know, make their case for reliability and validity in any way that they feel it demonstrates those characteristics. So, actually, we try not to be particularly directive and let the developers make the case themselves.

Kristi Mitchell: OK, thank you. I appreciate it.

In terms of parts to validity, so I don't believe that there are any admission – I don't think that if patients were inappropriately excluded or admissions excluded from the measure, I am still wondering if there is a table somewhere in the documentation that shows a variation across providers to highlight frequency in variation. Was that provided?

Susannah Bernheim: Yes. I'm searching my copy of the form you have, and if somebody else from the team is getting to it quicker than me, we do generally show the range of results at the provider level. And I'm just looking to see where that is.

Henry Ting: You mean the range of performance for this measure?

Kristi Mitchell: Yes.

Susannah Bernheim: Right, isn't that what you're asking for?

Kristi Mitchell: Yes.

Henry Ting: At the hospital level, Kristi, is 14.8 percent as a media and about 10 to 21 percent as the range of performance for the hospital level over 30 days or more.

Kristi Mitchell: OK, thank you. Thank you.

So, I can move on to feasibility.

So, for the most part, it seems like the measures are defined – the measures or rather the variables needed for that risk model are being found in the EHR. There were a couple of instances where the measure developer indicated that there were some low feasibility on, for example, the hospital's upper limit normal for component which is needed for risk adjustment, and that they – in talking with the – or through the EHR survey, it suggested that it could become feasible if manually captured outside the EHR system which would support calculation of the measure.

I mean, I wonder about the workflow issues in and around capturing that, but I don't know if you guys have tackled that or sought through exactly, you know, how that would work.

Susannah Bernheim: So, this is, again, Susannah Bernheim from the development team. So, what I'll say is these outcome measures, because the outcome itself – the mortality is not like going to be available within the hospital EHR by their nature, both because of the need to go to an outside data source for the outcome and because of the way that the models are calculated which is using data from all of the site simultaneously. And you've referenced this earlier, unlike some eCQMs which maybe able to be calculated on site as a hospital, these will always or at least initially for some time need to be calculated essentially in the same way that the claims and measures are currently.

And so, there is, you know, the actual workflow of how the data gets sent from the hospital EHR potentially with this upper limited normal included to CMS I don't think is worked out but it will – there's a lot of work right now at CMS around implementation of these measures and how to import data. And so, I think, you know, that's going to be part of what is the next stage for implementation of this measure.

Kristi Mitchell: OK, thank you.

And so, finally, to wrap this up, is around usability. So, the measure is currently not being used but it's likely to be included in future CMS Hospital Inpatient Quality Reporting Program.

And I think that's about it.

Reva Winkler: Liz, do you have anything?

Elizabeth Delong: Yes. This is Liz Delong. My – I'm a bit confused about the matching because according to what I read, only – now I can't find my notes. I think something like 25 percent of the org data didn't match and something like 53 percent of the CMS data didn't match. And I'm wondering what the impact. 25 percent or ERG patients weren't matched and that to troponin ratio between the match and the unmatched was somewhat different. I know troponin had a huge variability, but discounting that, I do wonder – I also wonder about troponin ratio in general because all a site would have to do would be to lower their upper limit of normal and benefit from the core because the ratio would go up.

Susannah Bernheim: So, let me respond to this too separately. Their – as you know, this was some differences in the matching between the Medicare data and the ERG data, and that likely has to do with, you know, slight differences in the capture of AMI on the registry side and on the CMS side is likely largely around fee for service Medicare versus Medicare advantage patients.

So, again, the measure, when it gets utilized, won't be using registry data. So, it will – similar to the current AMI measures to be identifying all of the Medicare patients who have a principal diagnosis of the AMI. So, those matching rates are really, really relevant for the model development we did in the ERG data and not so critical for the eventual use of the measure itself because then we'll be – we won't be using registry for that.

Elizabeth Delong: So, it would seem that it would be important once you adopt this – if this measure were adopted to see what the matching turns up. Presumably it would be closer to 200 percent.

Kristi Mitchell: So, on the registry side, they have a number of mechanisms to try to ensure that the registry are capturing all AMIs but it's – I don't – I think there will always be some mismatch between what, you know, a registry that's dependent on, you know, individual finding AMI within a hospital, find versus what is coded as a principal discharge diagnosis of AMI. So, I suspect there'll always be some mismatch there, but again, the registry is not

incorporated into the final use of the measure. The measure will just depend on identifying patients who have a principal discharge diagnosis of AMI much like most of the other measures that are currently in use for AMI patients. So, I'd expect that the cohort will mirror those other AMI measures when it gets put in use. I'm not sure I'm answering your question.

Elizabeth Delong: All I'm saying is that we're presuming that the two will be coincident and, I guess, because the hospital will be reporting from their EMR to CMS, the percent match will go up quite high. But it's still – it still seems it needs to be investigated.

Henry Ting: I would agree with Liz because we're using the action clinical database to do the risk adjustment, right? Those are the clinical variables that do the risk adjustment of likely that some of the – the expected mortality rate at 30 days for someone and that's what the importance of the clinical database adds is, the risk adjustment piece model. And we want to make sure that they're actually matched to the right patients to make sure that your risk adjustment model is appropriate.

I think that's maybe what Liz is asking about.

And the matching (inaudible) itself is that a unique patient identifier or are we talking about probabilistic matching?

Susannah Bernheim: So, I'm really glad you asked this question because I think we haven't been clear enough about the measure itself. So, I'll answer your second question first just briefly (and to say) not a unique patient identifier, it was deterministic matching with a number of variables that's being used and studied before with this registry with very high success rates.

But again, and I want to be careful to say this clearly because the ERG registry was used only as a tool for data to develop a risk model. When this measure is implemented, the registry will be completely separate from this measure. The plan for implementation of this measure is that we have identified variables that should be able to – be pulled out of any hospital's EHR regardless of involvement with the registry. They don't – they won't

necessarily overlap with the registry at all right now. The registry uses a completely different system for collecting these variables.

We were trying to just simply use a data source that had sufficient numbers to develop a risk model, but the measure itself doesn't use ERG. The measure itself will collect patients from any hospital EHR using printable discharge diagnosis and pulling directly from the electronic health record those risk adjustment variable.

So that we know …

Elizabeth Delong: So …

Susannah Bernheim: … matching with ERG in the measure when it is implemented.

Elizabeth Delong: So, basically …

Susannah Bernheim: Does that make sense?

Elizabeth Delong: … you're saying that the unique identifiers will be there. So, you don't have to do what you're calling deterministic matching but it seems as though it actually involves some level of probabilistic matching.

But then, as – was that Henry who said that you've developed this model based on the ERG data. And presumably, that's what you intend to use, the coefficients from that very model, is that correct?

Henry Ting: Yes, I think that's …

Susannah Bernheim: No. That's a great question. Sorry, go ahead. Henry, did you want to say something else before I answer?

Henry Ting: No, no, go ahead. I think Liz got – what she said was exactly right. I was (understanding) that ERG to develop the risk model.

Susannah Bernheim: Right. So, it's great. This is a really important opportunity to continue to clarify and it was – this was not a simple process. So, I'm glad that we can (assume) a little better.

So, the – when I talk about the risk model, I mean, primarily, the cohort, the variables used for risk adjustment and how well that model performs is what we were able to study using the registry data. But we are not – we don't set fixed coefficients, so every time this measure is run, it will be based on the – that run of the data. So, the coefficients aren't set from the risk model. All we did there was to determine that we had risk adjustment variables that could adequately capture severity in the hospitals to build a model on which the measure is based.

But when this measure is run, each hospital will send their cohort of patients with the risk adjustment variables, the first measure of all of those variables, and that will be linked to outcome data which will then be used to calculate, you know, across all hospitals with the coefficients that are calculated new for that cohort of patients.

May I pause just for one second to see if anybody else on the team wants to add anything to that so that I'm as clear as possible. (Bob), I don't know if you want to try to say anything further.

(Bob): Right, yes. This is (inaudible). Yes, I agree with – that I think there was a matching concern. I think Susannah explained well in terms of the fact that we just used the data to develop or to select the appropriate variables. And so, the eventual cohort will mirror exactly what the current cohort is give or take with the ICD-10 coming in.

The idea of using – the other point is a very good point that we developed, you know, which or we chose which variables to use based upon the action database which, you know …

Susannah Bernheim: Right.

(Bob): … may not match completely …

Susannah Bernheim: Yes.

(Bob): … with the CMS. But, as Susannah says, it's going to be mainly the, you know, the coefficients will change. So, in the future, once the electronic health record develops to the point where a new system could be put in, they could choose different variables.

Elizabeth Delong: That's – I think, you know, Henry nailed it that you're developing this not only on this cohort but with the assumption that these variables are going to be the most important, and that may not be the case if you don't have a bias subset.

Mary George: Yes, I guess – this is Mary George, and I was a little concerned or had a question about whether there were data elements in the ERG and that you would have liked to have had in the eMeasure but weren't available. The structured data elements was – were there any sorts of compromises on important data elements?

(Bob): I think …

Susannah Bernheim: So, that's a great question, Mary. Go ahead, (Bob).

(Bob): OK, sorry. Yes, I can deal with that. I think that that is a great question, and essentially we did come up with three that we would have liked. I mean, one was the electrocardiogram itself, how well it can, you know, whether it's an ST segment elevation MI or anything else that could be put on, but the electronic health record is just not at that level.

Other things that were brought up by clinical experts were heart failure on admission and cardiac arrest or shock on admission. And those just not – the data elements themselves even, you know, in the current definitions, it's so very difficult that we did not think that it was a feasible element to use. At this point, I'm unclear of how soon in the future will be able to be used.

Mary George: Is that because of …

Susannah Bernheim: And importantly, just to add …

Mary George: I was just going to ask if presence on admission is not available electronically yet.

(Bob): I'm sorry. What present on admission?

Mary George: You mentioned heart failure on admission.

(Bob): Well, there's frequent …

Mary George: That was the first …

(Bob): Right, but it did not meet our criteria for how well it's standardized across …

Mary George: OK, OK.

(Bob): … people, but heart failure on admission has been shown to have a very high variability how people code it. And we were very …

Susannah Bernheim: Importantly to add …

(Bob): … keep as objective as possible. Sorry.

Susannah Bernheim: And I just want – sorry, let me just – (Bob), let me just add to what you said. Again, this is Susannah, which is that in the case of the ST elevation MI question, the advantage of doing some of this work in a dataset that wasn't limited to what you could pull out of the EHR was that we were actually to test the model and understand how important that variable was because that would one that clinicians obviously felt might be a critical variable for the risk model.

And so, we actually ran the model both with and without ST elevation and it unfortunately made almost no difference at all to the performance of the measure. And so, we were able to feel comfortable bringing this measure forward without that variable knowing that in the future it may become available but then it's not – it's just not standard in EHRs right now.

Mary George: So, along those lines, I noted that we didn't see the correspondents between site specific predicted mortality and actually mortality. Did you do that? I mean, once you've developed the model, you have a prediction for each site for their expected mortality – of their – yes, their predicted mortality and their observed mortality.

Susannah Bernheim: Right. We often do do that for our measures and I'm actually pulling up our NQF measure now because I don't remember if that's in there.

I wanted to ask (Jeff), our team is having a very hard time getting on to the call and some of my support is not on the call. I don't know if you can check with the operator if there's a problem with the line.

Female: Operator?

Operator: Yes, ma'am?

Female: Yes, hi. I was wondering if you can open up all the lines just in case people want to speak. I know that we're having some difficulties getting some people on the line.

Operator: Yes, ma'am, the lines – all lines have been opened.

Susannah Bernheim: Operator, I don't think it's a matter of being open, I think they're just not getting picked up.

Operator: All lines are open and there's no one dialing in in the queue at this time.

Susannah Bernheim: OK. Well, I'll tell my team to try and dial in again.

OK. So, I now have our testing form opened. Yes.

Susannah Bernheim: So, it doesn't look like we did that. What we did do, as you saw, because we have a prior measure that is measuring the same thing, was to validate this measure comparing hospital scores on this compared to the current claims based measure that's used in public reporting and show that the scores for hospitals are very similar. I don't think we included the testing of observed to predicted for – at the hospital level.

Elizabeth Delong: I was just curious about the range.

So, I still wonder about using the troponin range. Maybe I've just never heard of using that. Is that a standard measure that's used? Troponin ratio?

Susannah Bernheim: It's important only because – right, it's important only because not all hospitals use the same troponin core. And so, to make something that's comparable across hospitals, you need to measure that troponin is relative for the normal range for that hospital.

Elizabeth Delong: Sure, I understand that. I just – it seems as though that is very dependent on how the hospital sets its range and that could alter the whole site's score.

(Bob): Yes. This is (Bob) (inaudible) again, and I guess the hospital could report an unusual upper limit of normal, but most of them are determined based on the particular assay that they're using.

I think that's going to be the big variability in general, but you're right. Specifically, somebody – depending on how it gets coded, somebody could put in anything they wanted.

Henry Ting: Hi, (Bob), this is Henry here. Thank you for being very helpful for your comments.

So, as I understand it then, we developed the risk model from the action registry, but then for the future reporting of this measure, it will be sort of these variables pull from the EHR as the risk adjustment model to look at – observe over expected mortality for 30 days after EMI.

Now, do we know that the, you know, that the data definitions for actions for heart failure or whatever these clinical variables are seemed to be important – are going to be the same as in these EHRs that are being automatically?

(Bob): No, I think that's a great question. And that comes into all eMeasures that are going to be developed whether to know about like this one is or just a retool will standardization across sites of the definitions. And that the – if you look at the five variables we chose and ended up being so important both statistically and clinically, they are very objective variables and they're pretty standard across so that you don't have a leeway in terms – so that's the reason we didn't use heart failure or even history of cerebral vascular disease or things like that or other clinical variables or other historical variables. It goes into age, blood pressure, heart rate, creatinine and troponin which are very

standardized across sites. And it should be very much translatable between the ERG registry and the electronic medical records.

Kristi Mitchell: Is that the same case for even defining the population to begin with, so, the patient population with MI? Do we (confidence) that it's standardized across EMR?

(Bob): Well, the idea being that they're going to be based upon the principle diagnosis through ICD-9 plus ICD-10. And as much as that's standardized, this measure would be standardized.

Reva Winkler: OK. So, this is Reva. Are there any other, you know, pertinent questions before we move on to another measure? Certainly, I think this measure will prompt a great deal of discussion at the in-person meeting and we're going to schedule extra time to be able to do that.

Are there any other specific questions for the developers particularly so they are prepared for that discussion?

Elizabeth Delong: I don't want to be the dead horse, but when they are prepared, I would like to see the range of troponin upper limits of normal and the variability that it possibly have used to that ratio measure.

Reva Winkler: OK.

(Bob): OK. We can get that.

Reva Winkler: OK. So, thanks to all the folks at (Yale). We really appreciate it and we look forward to further discussion about this measure at our in-person meeting.

And for the workgroup to proceed on the rest of our agenda, our next measure is 0133, we're moving on to measures the PCI mortality. This is the In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI. And George Philippides, you're the lead, and Mary George is the second.

George Philippides: Hey guys, can you hear me?

Reva Winkler: Yes, George.

Henry Ting: Yes.

George Philippides: OK, good. I'm calling from (inaudible). I just wanted to be sure.

So, our discussion sort of leads then nicely to this discussion because many of us – some issues do come up and I'll sort of get to that. And as you mentioned, this measure is basically, you know, 30 – I'm sorry, a hospital risk adjusted mortality rate using the NCDR CathPCI risk score that's been developed and validated over one million PCI procedures.

The one issue that many of our reviewers focused on was the issue of performance gap. There was a feeling that there was a small but significant performance gap. There is one area that I wanted to raise which is the gender disparity. And again, when I looked at the (inaudible) ratio for the gender disparities on page 49, it was but significant. And I'm wondering sort of moving forward will this be something that can be reevaluated such that as larger gaps arise, we can sort of work on that.

Reva Winkler: Do we have any measure developers on the line?

Frederick Masoudi: Yes. Fred Masoudi here. I'm not sure I understood the question. I'm sorry.

George Philippides: So, a gender gap, and that was small. Is there any sort of interim evaluation to see if the gaps increase? Can we then take a look back and see about whether or not they should be stratified in relation to these issues?

Frederick Masoudi: You mean look retrospectively from now?

George Philippides: Yes.

Frederick Masoudi: I mean, I guess, that's possible. It wasn't, you know, that sort of thing wasn't requested. The measure, you know, development within that fairly voluminous analysis that went into the measure development documents, but I suppose that that were viewed as a key priority, you know, that I think those data are obtainable.

George Philippides: OK.

And then moving forward to reliability testing, some of the same issues that were brought up in the prior discussion in regards to the measurement specification data was extracted from the action registry and not the EMR. And how appropriate that was? That was sort of discussed before, and I'm assuming that same thing that hold through here.

I think, Henry, you brought that (inaudible) for this one.

Frederick Masoudi: Yes, I mean, just to clarify the distinction between this measure and the last one. So, the last measure was eMeasure that was using action to validate the eMeasure. This is a separate measure that is constrained to the action there with the guidelines registry.

George Philippides: OK. All right. And then in regards to validity testing, one of the reviewers brought up the issue that certain (inaudible) or drop at it the data is not an element. And there was a question in (GFR), I think it was 25 percent of the cases or more can have – for that (inaudible).

Frederick Masoudi: I apologize. It may just be my connection but I wasn't able to hear the question, you're dropping in and out, and I apologize.

George Philippides: It's OK. So, one of the reviewers raised a concern that missing data elements was just (GFR) which, I guess, were not present in the fairly high percentage of the reports might had an impact or might represent a threat to validity.

Frederick Masoudi: Yes, in general, in the risk model, and I apologize, I'm struggling to fined the – my copy of the application here. But in general, and we can verify this for this particular model. But, yes, in the risk models, variables that are included are generally when missing at a relatively high proportion or, you know, use multiple (indication) techniques to identify those covariance. Again, I can very that for this particular model but that's generally approach that's been used in (MTR) risk models.

George Philippides: OK. And then in regard to feasibility, several of the reviewers express concern that, you know, (inaudible) or easily extractable and you can in EMR. But the data source is also included, the administrative database, for example, death. And in occasion, it might require ED or, you know, EMS records that might require manual extractions. So, I guess, the question was raised as to how easily it will be to extract this data for the EMR moving forward.

Frederick Masoudi: Well, this – if I'm not mistaken, this particular measure is not an EMR measure. This is, again, a registry measure. So, the data are all contained within the registry and inputted into the registry. And that's the mechanism whereby – so it isn't specifically, at this time, EMR extraction of any data.

George Philippides: OK. So, basically, it's participation in the registry that extract the data.

Frederick Masoudi: That's – but, yes, having the data within the registry, that's the source of data itself.

George Philippides: OK. And then in regards to usability, is this measure presumably being publicly reported?

Frederick Masoudi: Public reported, no. However, it is being use within the action – it's being used within the CathPCI registry itself in terms of feedback to specific – to sites with respect to their performance.

George Philippides: OK.

Kristi Mitchell: Fred, this is Kristi. I have a question about that. So, the measure is not being reported by private payers publicly?

Frederick Masoudi: There may be some circumstances where payers using this as part of reorganization programs, but it's not, to my knowledge, and I can again – I can verify this. It's not, to my knowledge, being used for public reporting per se of the actual results.

Kristi Mitchell: Thank you.

George Philippides: I think this ought to be raised by our reviewers.

Reva Winkler: OK. Mary George, you also were (inaudible) for this measure, any comments from you?

Mary George: So, I – more in terms of the variability for the rate of depth, maybe, Fred, if you could comment a little bit on how you're interpreting that range in terms of room for improvement.

Frederick Masoudi: Yes, I mean, again, I apologize that I don't have the materials right in front of me in terms of the range. And obviously, I think a lot of these issues around ranges are, you know, subjective. And obviously, also, with respect to outcome measures, it wouldn't be – it wouldn't necessarily be appropriate to expect a 0 percent – 0 percent performance for any of these measures on the one hand. And on the other hand, it's a little bit harder to answer that with respect to mortality itself because obviously that's an incredibly undesired consequence.

And so, one sort of imagine, there could be a circumstance where mortality rates were, say, 10 percent with relatively little variability around that, but that would suggest that there was global room for improvement and that's where – so this issue about variability around and mortality where it becomes a little bit tricky.

Mary George: Yes. And I know you there was a comment in there that there was a slight – a higher mortality in 2012 and 2011, and (I did so) if you had data go back from 2011 to see what that trend was that we say the last five years that might be something that would be helpful for us. You know …

Frederick Masoudi: I think we – yes, I think …

Mary George: I'm sorry, I was just going to say I'm very appreciative that you had very current data that's good to see.

Frederick Masoudi: Which I think points out, you know, again, one of the relative strengths of these is that the data is relatively contemporary.

Mary George: But, I guess you know, from my perspective, if we aren't seeing the trend where it's actually going up, and then that's over, say, a five-year period of time, that would be important for us to know in the committee.

Frederick Masoudi: Well, I believe, we should be able to get national – national rates going back further than two years. We should be able to supply those.

Reva Winkler: Anything else about this measure, any comments or questions from any of the other workgroup members?

All right, are you ready to move on to the next measures? The next one is 0535, and this is the 30-day all-cause risk-standardized mortality rates following PCI. And the next two measures are very similar. The first one, 0535, is for patients without STMIE and without cardiogenic shocks as compared to the other measure which we'll be with. But we'll start with 0535. And, Henry, I think this is yours.

Henry Ting: Yes, thanks.

So, this is the – it's sort of similar to the one we just reviewed except 30-day mortality is also similar to the (inaudible), the one that Kristi Mitchell and Liz Delong presented about hospital 30-day risk-standardized mortality for acute myocardial infarction eMeasure, but this is not an eMeasures, it's not for cute myocardial infarction, but it is similar concept for PCI looking at 30-day mortality after PCI using the CathPCI risk model for adjustment. So, all the same issues we discussed about risk model for adjustments and then trying to calculate observe to expected 30-day mortality rate after PCI.

I, you know, looking at the – regarding the evidence and the (regular) tools as that was actually done. And, I guess, a couple of comments with both Fred and (Bob) on phone is, we're using a risk model adjustment that includes all patient in the CathPCI which is age agnostic other than age within 18, I would think. But that's the same risk model we used then for CMS patients which are only over age 65. So, wondering if that is, you know, a similar (valid of elimination) that the factors that cause mortality in all patients of all ages are the same in those over age 65.

I think the performance gap, the range of performance is quite narrow. You know, it's probably a little broader than the inpatient mortality, but it's still quite narrow. The range of performance is 1 percent to 4 percent for the (best in the world's) performers. So, that's a range of (inaudible), you know, it seemed that the outcome (none of us launched) which is mortality, the one with that 4 percent doesn't mean to – a very small distribution of top versus worst performers.

Lets see. We answered this before was probably the term (holistic) matching to link the (inaudible) patient with CMS. And the only question I think I brought of was – before is excluding hospitals with less than 25 PCI cases which are in the database and probably the ones with probably the biggest problems on quality in terms of process measure, outcome measures and they're all being excluded from this measure.

And I think much of my comments that I had, I'd be happy to entertain questions from a colleague on the NQF committee.

Reva Winkler: Perhaps the measure developer can respond in some of Dr. Ting's comments?

Frederick Masoudi: Sure. With respect to the sample sizes, this is particularly important in terms providing robust estimates around an event like mortality, so, even more so important than perhaps among the process measures. And so, the sample size has to do with the – has to do with the ability to being able to provide reasonably robust testaments to the individual site level.

I – with respect to the issue of Medicare patients and young patients, I'd suppose there's always a possibility that there were age interactions with particular risk factors for mortality. I think once one get, you know, one can think about interactions on a, you know, wide range of scale. One could say, "Well, should there be a separate risk model for women or African-American men," and becomes, you know, difficult to, you know, interpret this, becomes progressively more difficult to interpret these models as one continues to – parts in terms of different demographic categoristics.

And can you think about some of the other issues you raised, Dr. Ting, maybe you can remind me.

Henry Ting: I just think, Fred, that if you – how do you deal with hospitals in less than 25 PCI per year because they exist? And I think (inaudible) they're probably seeing the most problematic ones in terms of quality, in terms of either ...

Frederick Masoudi: Well, yes, they may be, but again, in terms of the mortality outcome measures, the problem is that one needs to be able to generate. And I'm sure Dr. Delong can speak to this, relatively, were about testaments for these things. And so, providing observed expected ratios when the denominator is beyond a certain size becomes very challenging. So, although, I completely appreciate what you're saying which is, that you would loved to be able include hospitals of any size in a measure, it becomes a very important issue with respect to determining whether or not that the estimate itself is robust.

Henry Ting: So then it would be excluded from the risk model derivation but it would still be included in the reporting if they participate in – if this is probably reported, right?

Frederick Masoudi: No, they're not eligible for reporting beyond a certain size – below a certain size.

Henry Ting: And include in the risk adjustment model or the reporting.

Frederick Masoudi: Correct.

Henry Ting: Got it.

Frederick Masoudi: The other – just while we're on the topic, I have been – there's a minor correction in the comment to Ms. Mitchell previously. The PCI in-hospital mortality measure is being currently used in one circumstance for public reporting. It's in Wisconsin Collaborative that's using it for the purpose of public reporting. So, I don't want to be propagating any misinformation here.

Henry Ting: Fred, what do you think the reporting of all publicly for accountability in terms of both the inpatient mortality as well as the 30-day mortality for PCI cases and reporting both as separate performance measures?

Frederick Masoudi: Well, I mean, I think, you know, there are so many different ways that one could conceive of reporting programs where one of these – some of these measures might be useful for some types of hospitals and some of them maybe useful for others. And so, I would view them as potentially complementary to one another.

Reva Winkler: Comments on the measure from any of the other workgroup members?

Well, perhaps it would be useful to look at measure 0536 also because the constructs here are fairly similar but different patient populations in the risk model is somewhat different. So, Dr. Cleveland, do you want to maybe comment on measure 0536?

Joe Cleveland: Absolutely, yes. Joe Cleveland here.

So, as we've talked about, this is really a hospital risk-standardized 30-day all-cause mortality following PCI for patients with STEMI or cardiogenic shock. Again, the data source for this is that NCDR as well as some administrative claims who are capturing mortality through the – through CMS.

The – it's really I think the evidence is robust and this is – I think it meets the criteria one evidence for this outcome measure. There is a gap and a little broader gap in this patient population (inaudible) expected. So now, we're seeing mortality rates of 10.8 to about 14.4 percent. So, again, really, which probably does reflect the fact, again, these are patients with STEMI and/or cardiogenic shock.

There was a very nice appendix that looked at any type of disparities and there were none that could be identified on race or socioeconomic status again using dual-eligible patients. I think, obviously, the priority of this is high. This is very high severity, high cost type of patient population and intervention.

So, I think that within the reliability and validity test in the model, I was impressed that the C statistic was good, and therefore, I did not actually have concerns with that.

Again, I think the feasibility with this data source is relatively – this is something that is done or, in other words, they're easily extractable from the NCDR database that our hospitals report too.

With regards to usability, we've already touched on. I think that this measure is not currently publicly reported but as you really read through document, I think the intent is to, again, have this as NQF endorsed if it does become NQF endorsed or as a maintenance measure becomes endorsed, there would be plans to roll this out in terms of the data. So, I think the specifics are it is a different patient population.

I guess, the other thing would be, you know, addressing something that was also brought up a little bit with NFQ measure 0133 overall inpatient risk adjusted. You know, I think that there probably is value particularly for drilling down into this particular patient set. So, with regards to, you know, competing measures, I think that there is harmonization with the other measures and this is an important patient population.

Those are my thoughts.

Reva Winkler: Any other workgroup members? Kristi, you are second (person) on this one.

(Off-mike)

Reva Winkler: Kristi, are you there?

Kristi Mitchell: Yes, I'm sorry.

Reva Winkler: Well, I'm having a hard time hearing you.

Kristi Mitchell: Sorry. Can you hear me now?

Reva Winkler: Yes, much better.

Kristi Mitchell: OK, great. I just said, no, I really didn't actually have anything else to add. Thank you.

Reva Winkler: Anything from any other workgroup members?

I have one question in terms of feasibility. And that is, on the 30-day measure, identifying those post-hospitalization deaths and bringing them – collecting that data and bringing it into the measure and how that is done.

Frederick Masoudi: Yes, it's great question. Fred Masoudi, NCDR. This is done using claims data. So, basically using – go ahead.

(Laura): Sorry, Fred. This is (Laura) speaking. I think it might be helpful if the (percentage) of question could clarify whether you're talking about the testing data that was provided to support this measure application which is Dr. Masoudi is assessing this based on administrative claims data for the intended approach for reporting this back to hospital systematically.

Reva Winkler: Well, this Reva at NQF. I was going to ask the question. I'm just either one, actually, and as I look at the construct of the measure and the specifications, all of date is through the CathPCI registry. However, deaths post-hospitalization aren't collected there. So, somehow that data has to be obtained and I was wondering, you know, and potentially what the plan is for putting those datasets together.

(Laura Slatter): OK. Dr. Masoudi, this is (Laura Slatter). Would you like me to address the question?

Frederick Masoudi: Yes, that's fine, (Laura).

(Laura Slatter): So, the approach – so, you know, again, this measure is going through maintenance. And during that interim, we had not worked out the implementation approach. But at the time these measures were initially submitted, we had expressed the desire that they would be reported on a broader patient population and just the Medicare patient data that was use for testing to support the measures receiving NQF endorsement. We are in process currently of sending off the CathPCI registry records to the Centers for Disease Control and Prevention for linking up to the National Death Index. And then that's how we'll be augmenting these vital status data to be used for reporting of this measure.

Does that …

Reva Winkler: Are you saying you have patient IDs, social security numbers, and names?

(Laura Slatter): Correct. Well, and if you're familiar – so, if you're familiar with the National Death Index, you're not limited by virtue of only having social security number in order to be able to perform a match. It's a very granular approach that they use to give you probability of mortality matches to a dataset. But, in fact, we do collect social security numbers in our CathPCI registry. And we have a high percentage of valid social security numbers that are currently being provided. And – but we are not – we have other ways that the CDC will be able to attempt to match the data that's being sent off from our registry in order to obtained vital status information.

Female: So, it's deterministic matching?

(Laura Slatter): Correct.

Reva Winkler: So …

(Laura Slatter): So, again, so – I mean, in general, yes, there are about five classes of probability for matching the vital status information with CDC. Again, it's very granular but the majority of our records we anticipate will be a deterministic match off of the Social Security number.

Reva Winkler: But is this the same case as the previous measure once it become – will this become an eMeasure and you won't have to worry about matching?

(Laura Slatter): This is not designed as an eMeasure.

Female: But I thought we also determine that using EMR data to identify patients who died probably isn't going to work. You're going to need to match it to something.

(Laura Slatter): So, I can't speak to the other measure that's been submitted by (Yale's) team. I mean, they did talk about a centralized approach for reporting that measure. In this instance, these measures were designed to be reported as a centralized (reproach) through the registry program.

Does that answer the question?

Reva Winkler: Yes, thank you very much.

Are there any other questions or thoughts from the rest of the workgroup?

Since we're not quite or are in time, I hope we didn't cutoff discussion on any of the earlier measures. Was there anything from any of the discussion of the earlier measures that anyone feels they want to discuss further?

All right. Then we have few people listening in the call. This is an opportunity for public comment. From anybody who may be listening in to this discussion to offer any thoughts. Anybody want to offer any comments?

All right, I'm not hearing anything.

OK. For the members of the workgroup, is there anything else you want to add before we (inaudible) off? No? OK.

And to our measure developers, thank you very much for joining with us today. We really appreciate your input. As I think everyone is aware, the full committee will be meeting in April. I believe you all are aware of the time and the logistics. And so, at that point, we look forward to having further discussion on these measures. And thank you all very, very much for your participation.

Male: Thank you.

Male: Thank you.

END