NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
February 19, 2014
11:00 a.m. ET

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

(Evan): Hi everyone and welcome to the Cost and Resource Use Question and Answer Call Number Two. We appreciate you all joining us today. Hopefully we'll be able to – have some rich discussion about the three measures in front of you and get you prepared to submit your preliminary evaluations and then actually evaluate the measures that are in-person meeting in March.

And so, at this point, we'll go ahead and start off by seeing who we have on the call, we'll do a roll call of the committee, we have some TEP members joining us today to be able to answer questions about their evaluation and hopefully we have some measure developer representatives as well, who should be able to answer questions you have about the measures.

So starting with the committee we – do we have (Brenda Haspli)? OK, (Lisa Lapps)? (Arial Baywood)? (Larry Becker)?

(Larry Becker): I'm here.

(Evan): Great, thanks. (Marian Clark)? (Cheryl Danberg)? (Jennifer Ians Huff)? Nancy Garrett? (Andrea Galzer)? (Stanley Hawkburg)? Martin Marciniak? (Matthew McHugh)? (James Mason)? Jack Needleman? (Jean Nelson)?

(Jean Nelson): Here.

(Evan): Great, thanks, (Jin).

(Crosstalk)

(Evan): Yes, I know, we were – hopefully the people will join us. I know that we have few people in the first call. Maybe won't be dialing in for this one, but hopefully we'll get some individual on this call. Here we – (Janice Orlawski)? (Caroline Perrier)?

(Caroline Perrier): Yes.

(Evan): Great, thanks, (Caroline). (John Ratlif)? (Andy Ryan)? (Joe Stefanski)?

(Joe Stefanski): Here.

(Evan): Great. (Tom Singh)? (Lina Walker)? (Bill Weintraub)? (Herbert Long)? And (Dolores Yanagihara)?

All right, so we have a few committee members, hopefully we'll have more joining us. We'll now read off the TEP members and see who we have. (Sona Alkatib)? (Leslie Chow)? (Ted Gibbons)? (Jud Hollander)? And (Tom Cachi).

OK, operator all the lines open. There are some people told me they were dialing in.

Operator: One moment, I'll open the line. OK, all lines are open now.

(Evan): OK. I think (inaudible) that earlier, do we have – I'm sure – did I (re-off) some people's names who weren't able to speak earlier?

(Jim Nason): This is (Jim Nason), I'm also on the line.

(Evan): All right, great. Anybody else?

OK, we'll be – we'll monitor it as people join, I know that some people had mentioned they were joining. So, do we have representatives from NCQA?

(Ben Hamon): Yes, this is (Ben Hamon).

(Evan): Hi, (Ben), how you doing?

(Ben Hamon): Fine thanks.

(Evan): And from Yale.

Male: Yes, hi. From Yale, we have Susannah Bernheim, (Leslie Ott), (Emily Riley) and (Steven) (Inaudible).

(Evan): Great, thanks a lot for joining us.

Great, so we'll structure this call very similar to the first call we had, we'll open up discussions and the measure specifications on the type of evaluation. Really any questions you had as you've been reviewing the measure specifications or supporting documentation that we provided. We – as a reminder that the preliminary evaluations are due next Monday, February 24th and you – they can be submitted on SharePoint.

So really, the purpose of this call is to ask any clarifying questions as you complete those evaluations. So we have a webinar running right now that's show in the agenda as we speak, but we do have access to all the documents that we provided with you so we can bring things up as we come to them.

So, at this point, I'll open it up for the committee members to ask any questions of each other or the TEP members or of the measure developers.

(Joe Stefanski): This is (Joe Stefanski).

(Evan): Yes, hi (Joe), how you doing?

(Joe Stefanski): Pretty good. I want to follow up on comments or questions I had in the first call.

(Evan): OK.

(Joe Stefanski): And actually, it had to do with under the comments I made that was – bare in mind that my comment was made against the, you know, the knowledge and

the background that all are Yale folks and CMS measures developers are probably operating in their capacity and it's hard for them to communicate with the other related measure teams there. I did go back and I looked at the number 2158, their Medicare spending for beneficiary measure.

(Evan): Yes.

(Joe Stefanski): Generalized one that we – I believe passed on for endorsement. In (there), they were very clear about the period of time that the measure cost were going to be accumulated in less than three days before admit the 30 days after. And neither of the two measures that we have before us right now, I'm going to talk about the fact that the measures really go from three days before admission.

(Evan): OK.

(Joe Stefanski): And this is a – I can understand why that might be and of the data that they were looking at, that was 2008, 2009 was used for developing the measure. During that period of time, Medicare had a kind of a serious glitch in their claims processes and it required people submitting bill to make the admission date and the date of service is initiated equal. You couldn't do the – what was really required of the 76-hour rule.

So that every record that they look at, first date of service or start data service and the admission date would have been equal. But in practice, that probably was not the case. And it wasn't until 2011 that the CMS claims processing software was fixed or they could have a different first data service from admission date.

And, what this kind of point out to me, in fact, that the measure itself does not reflect those three days before admission, is that we probably need to make a change in our specs, the measure submission. Ask the measure developers to do a review of change, Medicare rules, regulations and problems that might affect data set (inaudible) data, data with the measures. That makes sense?

(Evan): Yes, I'll point that to developers that I know – our (2) Yale measures use Medicare data. I wonder if they have any comments on that.

Susannah Bernheim: Yes, hi, thank you. This is Susannah Bernheim. Can you hear me?

(Evan): Yes, actually, Susannah sorry, before you start, I've been receiving a bunch of e-mails here, I'm trying to monitor it as we have the call. Apparently there was some difficulty getting on the call today. So, I know we definitely have (Lisa Lapps) said she has joined. (Lisa), are you there?

(Lisa Lapps): Yes, I am, it was really not easy today.

(Evan): Yes, we – I'm sorry, I mean, this is a, you know, we're using the same dial and so I don't know what the – we'll talk with our vendor about this, but hopefully we don't run to this in the future, but I know (inaudible).

(Joe Stefanski): There were a number of problems yesterday with the socioeconomic risk adjustment group.

(Evan): OK.

(Joe Stefanski): Not being able to get in. So there's something going on with your vendor.

(Evan): Yes, no, we – I apologize profusely, this is – it's not acceptable, so we will – we'll make sure we get that address. But I really apologize for today. I know that (Ted Gibbons) member of the TEP, we're you able to get through? All right, send me an e-mail. (Arial Baywood)?

(Arial Baywood): Yes, I'm here. The invite also didn't have a number on it, the current – I mean, I found it on the old one.

(Evan): OK. So – OK, we'll try to get this address, so really, I apologize. I'll send another e-mail right now just to make sure everybody has the invite. But so at this point, I'll go to Susannah, if you want to address the previous question.

Susannah Bernheim: Sure, and just so some people just joined, the question was about the three-day rule. And I'll explain how we thought about it and we also have our analyst on the phone, when I'm done, I'll let her weigh in to see if I forgot anything crucial.

So we do look pretty carefully at the Medicare rules and we update the measure every year based on those. Our understanding was that related payments in the three days leading up to the admission aren't procreated into the DRG payment. And our goal would be only to include those related payments. So we did not look specifically for claims outside of the index admission claims during that three-day window, because they were paid for separately, they are thought to be unrelated and that the related claims are bundled into the DRG.

And since we're using the DRG, we should be capturing that. And so we consciously made a decision based on our understanding of the claims and our measure therefore, should be capturing the related payments on those first three days.

(Leslie), do you want to add anything to what I've said, I may have oversimplified it a little.

(Joe Stefanski): Let me butt in here for a moment. It's not so much to – I agree that now in going forward, it will indeed capture those related charges in the three-day window. What I would like to see is if we're going to talk about that directly and say measure 2158 and make it very clear that we're including those in the measure, we need to have consistency on the way these major specs are written.

Susannah Bernheim: Right. I'm not sure, you know, in all honesty that the measures are consistent. Our reading of the MSPV measure was that even potentially unrelated claims are included that all claims that come in during that window, so it's actually handled differently in these two measures.

(Joe Stefanski): All right. And I guess that it needs to be made very explicit or we'll loss track of this later.

Susannah Bernheim: OK, so we will – I think some – the Yale and what we can do is try to make sure that our language around the assumptions about what we're capturing gets written out, I know it is in our technical report, so we can look at that language to be clear about what we're doing.

(Joe Stefanski): And one of the things that I'm raising and I'm just using this as an example, is the issue of changes in Medicare regulation. And in this case, I'll give you an example of a readmission, the CMS readmission measure. We think it really got lost along the way.

At one point, CMS contract with a QIO, you know, the Quality Improvement Organizations in a particular state to do with some programming for them on the readmission measure. And that – those – that software which is fast code essentially get used all over the country in the QIOs.

But they forgot was that regulations change. And so, when we started to see in July of 2011, when the software has corrected its Medicare and that the starting date of service could be different than the admission date, the software was written originally assuming that the first date of service was equal to the admission date. And therefore, they were miscalculating the admissions, not by a lot, but it was adding in quite a bit of noise to the measurement and we're trying to eliminate as much noise out of these measures as we can.

It is very easy for people to miss changes in regulation or problems that CMS software with called issues down the line and add those to measures.

Susannah Bernheim: So obviously, we can't come it all on to have CMS programs. I will say that our team reevaluates the regulation for all of the settings that we're pulling in on annual basis because it obviously does change year to year. And so, you know, each year we'll be updated with, you know, (data) based on them.

(Evan): Great, do we have more questions about the measure specifications? First, let me ask and I reset up the dialing information. So people try to dial in. Is anybody been able to join us?

Nancy Garrett: Yes, this Nancy Garrett from (inaudible) …

(Ted Gibbons): This is (Ted Gibbons), I've been able to join.

(Evan): Great, thank you. We apologize profusely, we'll get this fixed, I know that mostly have a problem with their vendors.

(Herbert Long): And this is (Herbert Long) from (Orica), I joined as well.

(Evan): Great, thank you.

(Jennifer Ians Huff): This is (Jennifer Ians Huff) with (PDPH).

(Evan): Great, thank you.

Jack Needleman: And this Jack Needleman, I joined late, sorry.

(Evan): I know people joining late and people couldn't join, so I, you know, I hope – I apologize again. I just want to remind everybody, we are streaming through the web, so if you are able to turn your computer's speakers down, that might eliminate some of the feedback we get through the – to the phone, but that's great. So for those …

Female: Sorry to interrupt (Evan), is there a web URL as well or is it just for (inaudible)? Is anything (beyond) web?

(Evan): There's a web – I send – to send it out around again, it was included in the agenda and I send it around again in the e-mail.

Female: OK.

(Evan): I'm really just – I'm just showing the agenda right now on the web link, but we have it set up so that if we need to show a certain specifications or any of our documentation, we can do that through the web – screen sharing.

Female: Thank you.

(Evan): But those of you who just joined, I'll go ahead and give our preamble again just –I know that we started off kind of slow on this call.

Again, the purpose of this call really is a question and answer. We're going to structure and similar to the first call, so – and there was kind of an either or for

these two calls. The difference with the second call was that we invited members of our technical expert panel to be available to answer questions about their evaluations that was included in the committee packets.

I updated them with track changes after their evaluation. So really, the purpose of this call, again, is to ask questions of each other, of the TEP, to measure developers, both measure developers are represented on this call or of NQF staff about the evaluation process.

So, this really is to prepare you to submit your preliminary evaluations which are due on February 24th and then for our in-person meeting which is March 4th and 5th.

So, with that, we can go ahead and continue the questions.

Do we have any questions?

(Joe Stefanski): (Inaudible) talk to the group.

(Evan): (Joe), you've been kind of carrying the mantel.

(Joe Stefanski): I listened in on the TEP call and I thought that was …

(Evan): Yes.

(Joe Stefanski): … pretty straightforward.

(Evan): OK.

(Joe Stefanski): And partially, because all of these clinical definitions have been examined by multiple groups because of multiple measures trying to use the same definitions, it was pretty good. And the few questions that have to be answered. And when I looked at the NQF measure which I like the first time around. And last – on the last call, we had the NQF representative give us kind of a quick summary of the changes in the measure from the first time around because someone is …

(Evan): And (Joe), just to stop you to clarify, we're talking about the NCQA measure.

(Joe Stefanski): I'm sorry not NQF, NCQA measure. I'm sorry.

(Evan): Yes. Just so everybody is on the same page.

(Joe Stefanski): Yes. It's, you know, I like – like I said, I liked it the first time around, I just wanted to know what the changes were in the measure so I could evaluate it against that and, you know, I thought that was fine. And these other two measures, the CHF and the AMI are really to meet pretty straightforward. So, I, you know, I just don't have a lot of questions.

(Caroline Perrier): So (Evan), this is (Caroline Perrier). I guess I have question on the NCQA (our new) measure, it seems that that from the TEP comments, there was some concern around the risk adjustment model. Could you just talk through, or could somebody just talk through that to – because I was just talking up a little discomfort on behalf of the TEP. Now this is – obviously, this is up for – it's a maintenance deal. But it's obvious it's been in placed for a while, but I'm wondering how this concerns relate to moving forward with this measure.

(Evan): Well at this point, if I call on the measure developers from NCQA, I know that one of the things that we provided along with the report, they provided us with the document report on the (HCC) models. And so, I guess I'll let (Ben) talk a little bit about that.

(Ben Hamon): The TEP members request was they – they were asking if we had done a reassessment of the risk adjustment model since the implementation. And we had not – primarily because, you know, we rely on CMS to maintain the CMS (HCC) model that we use for risk adjustment but also, because NCQA does not receive patient level data for the (RE) measures, only in aggregate.

So, it is very difficult for us to, you know, to sort of reevaluate this model. You know, we did it initially in testing when we first determine the appropriate model but, you know, the annual data we get from each plans only in the aggregate until we could create theoretical retest models but without getting a patient level data again, it would be very hard for us to do that.

And so, I don't know if that explains the question – our response if you wanted some more information about that.

(Evan): (Caroline), does that answer your question?

(Caroline Perrier): So, let me just, again, ask, you know, OK, yes, that's answers the question. So you don't have patient level data in order to use – or for this particular measure but with the other two, are we using patient level data because we're using a different data set?

(Evan): Yale, you want to speak to that?

Female: Sure, yes. We have risk adjustment data on all of the individual patients including the measure from the claim.

(Caroline Perrier): OK.

Male: I mean, the underlying questions whether a risk adjustment model develop using Medicare patients, Medicare fee-for-service patients is appropriate for commercial products is going to be included in the NCQA.

(Ben Hamon): Right. And so our initial testing when we determine the appropriateness of model, we did use a large database that included commercial patients and it was found to be, you know, appropriate. And I believe that is included in the original testing information …

Male: Yes.

(Ben Hamon): … so the response is we just didn't retest that again and again on an annual basis since (inaudible) initial test, it's done, I think, in 2009.

Male: Right. I remember our discussions about that model back the first time here. But I think there's only two or three of us on this committee that will – on the present committee that we're on that committee looked at that measure.

(Caroline Perrier): OK, thanks for – thank you. That is helpful.

And then another question I had and I'd asked this of the group, again, because I just don't know enough, you know, the big question in my mind as it relates to consumers or people that might be looking at this information to help inform some of their decisions about where to seek care, do we feel that these measures give us enough variability so that you can really discern quality differences when you look at these numbers?

And I'd just like to hear other's opinions on this.

Male: Well, I can speak for the NCQA measure that we see in HEDIS. And in fact, particularly for this measure, we do see – because we report this measure in HEDIS using quality measures and the (RE) measure in conjunction which you can see on the sample score report. We do see quite a large bit of variability between different health plans both on the quality and the resources (domain).

(Cheryl Danberg): Can I ask a follow up question to the NCQA measure because I don't believe I saw this in the documentation, this is (Cheryl Danberg).

One of the things that I think I was looking for because these measures have been in play was some sense that organizations that had been measured on them and where they've been in use that this had led to improvements overtime.

Are there data that suggest that plans have been able to act on the information?

(Ben Hamon): These are – because of the nature of the measures, these are very hard measures to actually trend and the – for the – in the resource use domain for HIDES, we're actually finding that the plans themselves are doing some internal analysis, you know, based on like I said, they have the access of the patient level data and so they can sort of deconstruct the measure results, based on the expected that we provide them from the calculations using our indirect standardization.

We're actually finding most of the – the greatest use for these measures are (inaudible). In the business, the employer sectors and the benefit designs

where people can actually look to see, you know, the states are very interested in the level of resources used because everyone seems to be interested into the value equation and the efficiency being offered by this health plans particularly with the quality.

So, we're hearing stories and plans are doing a lot with the, you know, with the information that NCQA provide as part of reporting for this. Like I've said, we provide each plan with individual expected ratios along with their regional and national benchmarks and they can intake that and do their own internal analysis to identify areas.

But, we really can't trend the measures year to year. So, you can't really watch a plan change its position year to year because of most of the indirect standardization is based on the data submitted by all the plans for that year. There's also adjustments to the standard pricing that's performed every year. And so, it's very difficult without sort of holding a number of things constant artificially to watch trending changes with plans year to year.

(Cheryl Danberg): Yes. I mean, I appreciate all that. But I guess I'm trying to understand because the things like other measures that (depart) in this larger performance measurement dash forward. You're able to assess whether improvement has occurred overtime and I'm just – I'm sort of scratching my head as to whether there's a companion metric that – or set of metrics that would allow someone to know that by plans focusing on this measure and maybe doing drilldown analysis and taking some actions that this has resulted in some shift in something.

And I guess the question is sort of how to connect the dots here. And – so that's what I was struggling with when I review the measure.

(Ben Hamon): Right. We have not through any formal mechanisms receive any information about plans, you know, receiving major changes and we do do annual analysis. Again, at the higher level looking for correlations between the results in a different plans but we don't dependably individual plan basis. We tend to look more at the regional results to look for correlations between the different components of the measure.

Just because there is a huge volume data here and I get it, people are more understood in the – they focused in on very specific aspects of these measures. Some people are very interested in the inpatient facility side, some people are very interested in the pharmacy. Some people are very interested in the outpatient side and so on and so forth.

And so, they tend to sort of drill in on the various aspects.

Nancy Garrett: And this is Nancy Garrett. Just to follow up to that about the use of the measure. Has it been – have you seen use of – at the provider level? So, health plans calculating the use measures for providers in using it to work with them on performance improvement.

(Ben Hamon): There was a pilot of provider groups in California that actually (Dolores), you probably can speak to it, I don't know if she's on the call or not.

It really – because of the volume of data that's required for the measures in the attribution level is really sort of, you know, at the plane level that's where the endorsement was for this measure. You know, the minimum sample size for this is 250 patients, which is often times even difficult for a plan to come up with especially not in the cardiovascular conditions so much but for other resources measure.

Male: What is the number per thousand? What is that translate to in a commercial basis serve for the Medicare population? Could a large …

(Ben Hamon): I didn't understand the question.

Male: Could a large ACO have the 250 members, if you had a big provider organization, or would you really need a …

(Ben Hamon): Yes, I think that's certainly is possible. Particularly like I've said for the cardiovascular and diabetes measures, the sample size isn't generally an issue for the larger entities even the provider groups (inaudible) that.

For some of the other measures like COPD and asthma which are other resources measures that are coming, the pulmonary call. We occasionally find

sample size issues because the plans don't, you know, don't meet that simple size, the minimum sample size.

Male: So …

(Crosstalk)

Male: Have you seen any organizations though incorporate this then into some sort of shared dating is a risk model to the original – can answer the original question?

(Ben Hamon): I've heard rumors. I have not – I don't have any verification of it – no.

Jack Needleman: Oh, this is Jack Needleman. Question on the CMS measures, no part D data. How much does the essence of information about outpatient drug use affect the consistency of the measure?

Susannah Bernheim: So this is Susannah Bernheim from the Yale team. It's a great question, as I'm sure you know, if not all beneficiaries have part D. And unlike parts A and D and which have pretty high rate and among people who are in Medicare, which is very high among the over 60,000 population a very high percentage of part D. So we are able to capture a really good and representative population and know that we have their full Medicare payments. But part D there is just less penetration and it's more inconsistent. And so it's very hard to trust if you use the part D that you're going to get kind of fair comparison across.

So the truth is we don't know the answer to that question because we haven't dug in for the part D data. We do have it and thought about trying to do some validation work.

You know, I suspect given the overall cost of the inpatient stay and the outpatient, you know, the costly outpatient pieces that happened that it's not going to dramatically change what we see. But right now, we just don't think that we got good enough data to include it. It's something we'll look forward in the future.

Jack Needleman: Right, well, just trying to think about, you know, if it – if the drug therapy to these conditions are fairly consistent and consistently applied, then the absence of the data doesn't make a lot of difference.

If there is a lot of variation in drug therapies and that spills over into use of other services because some drug therapies are less effective than others, some require more physician monitoring, then we're picking up variation associated with differences in drug treatment without having any information about that.

Susannah Bernheim: Right, no, I mean, I …

Jack Needleman: I'm not a physician so I don't know which of those is truer.

Susannah Bernheim: I mean, in general, except for, you know, some specialized situations like chemotherapy, there's not that many medications given on an outpatient basis that are going to be costly enough in a 30-day window to sort of overwhelm the patterns we're seeing especially in these particular cohorts.

But – I mean, I think it's something that we could try to look into a little bit more if we found enough reason where there was enough penetration that we could trust our findings. But, there wasn't a valid way to incorporate it. So, I think it's really unlikely that it would change the pattern. But I think it's probably something worth (as) continuing to try to look into in the future.

(Ted Gibbons): (Ted Gibbons), I have a question of the NCQA representative.

(Evan): Yes, go ahead or Ben, are you still there?

(Ben Hamon): Yes, I'm still here.

(Evan): OK, good.

(Ted Gibbons): This is (Ted Gibbons). I was wondering, even though you don't have patient level data and it sounds from your description of the current utility of the data that you're depending on some of the users of the information such as health plans and states to look into their patient level data to make changes in quality improvements. Before even getting there, what was the objective of your measure development? What did you hope to gain by information on a broader level? Were you looking at reducing the cost of unnecessary testing or what was the broader picture of why you developed this measure? And in sharing it with the health plans?

(Ben Hamon): Well, you know, again, I think the biggest issue was we were tasked to and this, you know, quite a few years ago to really develop measures of healthcare value. And that went under the guise of efficiency and resource use and really trying to understand what a measure would look like that would really get at that. You know, the value and the quality for the service that's being delivered by health plan in the interest of, you know, carrying plans and ranking plans as far as what they're offering to their customers. And, you know, they looked at several different approaches and what you see is effectively what we have come up with.

(Larry Becker): So, this is (Larry). So I think it's obvious to everybody else but why the decision not to use patient level data?

(Ben Hamon): Well NCQA – for HEDIS reporting did not ever receive patient level data. You know, we test each patient level – (use) patient level data but HEDIS reporting is aggregate level data that is verified by HEDIS auditors to go ahead and actually, you know, verify the data if it's accurate and being submitted.

We only receive aggregate results so that, you know, when we test and develop a measure, we use the patient level data to test and develop. But the annual HEDIS reporting only the plans only report to us the, you know, the total patient, the patient cohorts for our (U) or just the rates for the (inaudible) measures.

(Larry Becker): So is it your contention that if you actually use the patient level data, the results would be the same?

(Ben Hamon): Well, the results are generated from patient level data. It's just that we don't receive it from the plans.

(Larry Becker): And you're saying how you get it? And what I'm asking is – the question is did you test the validity and does your data suggest that if you did it directly with patient data, you get the same result?

(Ben Hamon): Yes, because when we tested the measure and developed a measure, we used patient level data and we used a very large database. It's comprised of about 60 to 80 large health plans that had several million members – member months over a several year period, so.

(Larry Becker): OK, all right.

(Ben Hamon): The measure reliability and validity was test using patient level data. But again, when we just, you know, I said the annual reporting that we get is not patient level.

(Larry Becker): OK.

(Ben Hamon): So I would certainly assume that this, you know, measure still hold up even if we were to retest the – something that we did a couple of years ago.

(Larry Becker): Could you tell us more about the auditing process for making sure that the data that goes into the measures are complete? I'm assuming as part of the auditing process is making sure the measures are constructed properly. But I'm more interested in the data.

(Ben Hamon): So the auditors are certified by NCQA. And a major part of their job is source verification also to make sure that the calculations are being accurately performed. In addition, they also make sure that the plans are following the standard structure of measures, the accreditation of the health plans that report the used measures have a large volume of structural measures which include, you know, the allowable data sources, the types of data to be allowed to be

reported and so on, and so these are trained individuals who have very sophisticated methods for verifying a data that's being aggregated and reported to us.

(Larry Becker): OK, so, you know, we were trying out the part D data just not being available for the Medicare folks consistently. But many of your health plans carved out both the drug benefits and mental health benefits. And in the last rounds of reviewing these measures, basically, you said the auditing process assures of the plans have gotten that data back from the carved out firms to incorporate into the measure. And I'm just wondering if you can speak a little more about that aspect of the issue.

Basically, took you at your word last time and said yes to the measure. And the folks who had measured this didn't have those pulling back, didn't get approved. So this is a key issue.

(Ben Hamon): So, the measure specifically will require the type of benefits that is required in order to report the measure. So, you know, for many HEDIS measures, it's primary just medical; for many of the medication based measures, that's medical and pharmacy.

And in that case, it is responsible – the plan's responsibility to access that pharmacy data in order to report the measure because I can't just say, you know, well, we don't have the pharmacy data and therefore our rate is this because, you know, and it would basically a false rate.

For the relative (inaudible), I believe we do not require the pharmacy benefit. They offered – only the medical benefit is offered. However, that the, you know, there are fairly distinct pharmacy categories in this, and a plan – a plan result that's going to be severely adversely affected by not having a data and they must, you know, again the auditors are the ones who are responsible for ensuring the planner actually using appropriate pharmacy data to report the measure and not just, you know, using a lot of blank fields to create the false rate of pharmacy utilization.

It's a plan – if a plan is reporting pharmacy data through the (RU), the auditors are verifying but, in fact, they are using pharmacy – they have access, whether

to carved our or not, you know, it's the plan's responsibility to appropriately report that data to us.

(Larry Becker): OK. And same thing is true for the mental health carve outs?

(Ben Hamon): Right. So, if the mental health benefit is required in the measure, again, which is not, you know, only the medical benefit is required for this measure that is – if that's the case, yes. The auditors are the ones who are primarily responsible for verifying that – that the plan is accessing the appropriate data even if it is carved out to report the measure and not just reporting, you know, the absence of and making themselves look artificially and, you know, ahead of the curve as far as resources for that specific population just because of the absence of the data.

(Larry Becker): Did you say mental health benefits are not required to be included in 1558?

(Ben Hamon): No, the only thing that's required for 1558 is the medical benefit. The other services are part of the standard pricing table so they are included, you know, is that an optional for patients but their, you know, like I've said, the only – that required benefit is medical.

(Larry Becker): OK.

(Ben Hamon): And we don't – and, you know, we do have discreet service categories for the pharmacy. We don't really for the mental health and to research as measures, not this time.

So plans report, you know, their pharmacy utilization needs and the (inaudible) resources. But we don't have, you know, a service category for mental health outside of it, you know, what are included in the outpatients and inpatients services fair price.

(Ted Gibbons): This is (Ted Gibbons) again. Can I ask a question about how information is shared among plans that used your data, for instance, is there a forum for individual plans or communities to share their findings based on patient level data such that they did institute quality improvement plan, for instance.

If there seems to have been an unusually high utilization of cardiac catherization for readmission after a myocardial infarction or a (stab) displaced where it may not have been necessary and that was – it was gleaned from patient level data. Is there a way, for instance, to share the methodology by which that conclusion was drawn among users of your information?

(Ben Hamon): So, we have certain restrictions for the reuse of HEDIS public reporting results, you know, and it's published to our quality compass module.

I don't think we restrict the individual plans. If they do their own internal analysis and they wish to share that information with other – with outside people, you know, because, again, that's something that's sort of outside of our control …

(Ted Gibbons): Yes.

(Ben Hamon): … it's beyond the measure reporting program. You know, I don't – I think it'd very interesting to us and I love to see it. But I don't think, you know, we don't either offer venue or offer any restrictions on their, you know, what their – what they can do and what they can share with various people outside of like I said just the official HEDIS reporting result is somewhat restricted. They're allowed to share but under certain terms.

(Ted Gibbons): Sure. So you're not sponsoring any way for them to communicate these QI improvements. But …

(Crosstalk)

(Ben Hamon): No. We make the data sets available so, you know, you are allowed to access the complete data set of all the, you know, the plan results for anyone reporting HEDIS, that's very popular with the researchers.

We don't currently have, you know, outside of some educational programs that we've offered over this last few years as far as, you know, different types of opportunities that we do to the HEDIS update conferences and, you know, we supported that for quite some time.

But as far as offering individual plans in a specific resources – at this time, we don't really, you know, in the early days we did offer plans a lot of assistance reporting to help, we were finding most of the errors where calculation errors. And so we did do a lot of, you know, we offered a lot of technical assistance to plans and help them get their – get the reporting right.

But most plans now have figured it out and don't require that assistance anymore. So, much of our education campaign now are around talking to other stakeholders. I guess on a business communities and others to help them understand the results.

(Ted Gibbons): Right. So it still makes me wonder how you were communicating your ideas about what you considered the equality profile to be.

(Ben Hamon): Well, outside of the quality compass module program where we have, you know, all the plan results in the health plan ranking which these – some of these results are being used for now. You know, again, we offer assistance when it's requested but we don't have any (fault layers) or any formal process right now for reaching out the plan for this time.

(Ted Gibbons): OK, thank you.

(Cheryl Danberg): This is (Cheryl Danberg). I had a question, because I was struggling with the reliability testing that was done and I tend to think about reliability testing as being able to discriminate one plan's performance from another. But it seems like that wasn't necessarily the intent of the measure that you're really looking to look for differences, either above or below (ones) and not to compare providers against each other.

Is that the correct interpretation?

(Ben Hamon): Much of our early reliability testing was, you know, because these are claims-based measures using large data sets, was really in the accuracy of the plan results. And so looking again, looking for calculation errors, looking for the comparability results and so most data analysis produced lying in that. Again, these are really, you know, these were thought about as plan level measures, they were not really thought about as individual provider level measures.

So, I don't know if that answers your question.

(Cheryl Danberg): Yes. But I think the intent is not necessarily to compare one plan to the next, right? Or is it? Because – again, what I was trying to figure out is, you know, what size denominator and how much variation, you know, between and within you need to see to get a reliable signal.

(Ben Hamon): Right. So it is actually a fact that, you know, it is a plan comparison tool where it gets a little fuzzy is when two plans are pretty close to each other and it's a matter of, you know, because we use so much risk adjustments indirect standardization, you know, whether those differences of plans are very close to each other or actually truly significantly different in either the resource user the quality dimension.

But as I've said, you know, there actually is rather large variation in playing performance, you know, at the regional level on both dimensions of our value equations in our quality and resource use. So it's rather surprising to see, you know, either at the same level of quality, the level of resources used to obtain that and really goes right across our scale.

And so – so it is a plan comparison idea and then, you know, it was based as, again, that sort of that the intent was to try to identify the value of services being provided for certain level of resources used. And, you know, again, but the specifics technology have limited, you know, how advance we can do – or actual display as far as, you know, presenting plans that are basically right around the mean which is the – so, you know, the reason we get the quadrant method is because I think it gives the purchasers a really good idea of, you know, plans that are achieving high quality at a relatively low resource use versus the ones that (teaches) low quality to very high resource use.

And so, we're, you know, we're very comfortable in the identification of those plans are the most quadrants. But again, you know, two plans are hovering right around the mean, you know, how different those are, if they're pretty close to each other is a little bit more challenging.

(Cheryl Danberg): Right. Thank you for that answer.

(Evan): Do we have more questions for the developers, for TEP members, for each other?

Nancy Garrett: This is Nancy Garrett. I just have a question for NCQA about the quality compass are used table that we have on the SharePoint site, it's called a plan B tilt table. Let me see if I can get the name of the document. It's called F10 sample score report.

(Ben Hamon): OK.

Nancy Garrett: Are you familiar with what I'm talking about?

Male: I'll pull it up right now. But it was submitted with the – yes, the measure submission.

Nancy Garrett: Yes. So I was just wondering if the NCQA person would mind walking us through the table because I'm not quite – or haven't interpreted it.

(Ben Hamon): OK. So, there are a number of components of the different (RE) measures. And in our quality compass presentation, you know, when we present the plan results back and I'll make them available for public reporting. We like to present both the detail level and the higher level like the quadrant view. You know, as I've said, the (RE) measure are broken into several components. We have the quality component which is a quality component index – composite index. You're looking at it at the top – the top row if you look at the table version here.

The total medical which is the roll up of a number of different components as you can see that for the further downward, there's the inpatient facility, procedure and surgery E&M, different components and that roll up the total medical.

Those are the two which we then present on the quadrant graph. So you're looking at the quality composite against the total medical index, genetic quadrant which is below. That's what's, you know, again, that idea of, you

know, what plans are high quality, low resource use versus low quality high resource use, et cetera and so on and so forth.

The other component and, again, you know, these are all normalized to the means so we look at plans that are over 1.0 as slightly higher, lower than 1.0 is slightly lower than the means. And again, we don't necessarily judge if that higher is better or lower is better for each individual components, you know. For example, I always like to use the idea that, you know, higher – for cardiovascular conditions perhaps higher outpatient E&M and higher pharmacy might significantly reduce, you know, put you below on the, you know, much higher quality of care and put you much lower of the inpatient utilization because you're doing – you're managing your patients in the outpatient setting.

And so, again, I don't – we don't judge on the – you just present the data. And so what you're seeing here is at the national level which is effectively all the regions the HHS rolled up in the regional level which is where that the region in which that plan is located, you know, what their results are.

And so, you know, we do that in detail. We also do that like I've said at the quadrant level which is kind of the grouping of the total medical and the overall quality composite. Some people like them to hire detail level and others like the sort of the more general.

Nancy Garrett: OK, so the total medical index is the actual measure?

(Ben Hamon): The total medical use – the total medical is the roll up of E&M procedure and surgery inpatient facility together. It also actually includes two new categories which I don't think made it to this one which is diagnostic lab and diagnostic imaging. So that's the total resources used for all of those components. And those are all the TNPMs, you know, standard price services that are included as part of the measurement strategy.

So the total is just the roll of all of these different service categories.

Nancy Garrett: So the total medical is actually the measure?

(Ben Hamon): Right, there are additional components of a measure. We do have some frequency of service components, the inpatient discharge and the ED discharge index is which your PMPY. We also have some other components, frequency of selective procedures such as carotids, and (EVTC) procedures for the patients who fall in this category. Those are not part of the total medical. But those are good comparison tools as far as, you know, what the frequency of services for patients in this category who are getting these services on a plan by plan basis.

Nancy Garrett: OK, so I think I'm struggling with understanding why the total medical index is under one. Wouldn't it be 1.0 on the national level? Isn't that what we're indexing to?

(Ben Hamon): Well we're indexing to all the plans that submit that the national level is basically the summation to all the – of all the – this plan in comparison at the national level. So, if the plan was, you know, if it was – if it was offering the same level of resources as a similar plan with a similar member base, you know, as to any other plan who had a similar size plan, a similar member based, it would be one. But again, in this case …

Nancy Garrett: Oh, we're looking at one specific plan here.

Male: Yes.

(Ben Hamon): Yes, it is a specific plan result.

Nancy Garrett: I didn't mention it. Got it. This is an example, OK.

(Ben Hamon): Right, yes. Yes, we just picked this, that's why you should see the blacked out section is the plan name and the sub ID.

Nancy Garrett: Got it.

(Ben Hamon): The blacked out there, it was just some random plan we picked.

Nancy Garrett: OK. That's helpful.

(Ben Hamon): It looks like there's plan because if you look at the scatter plots, the little red dot is the actual plan. It looks like it actually pretty much right on the mean for both quality and the resource uses, I mean, that's why we picked that, I don't know. That's how the plan would appear in the end results in our quality compass module.

Nancy Garrett: And again, I'm sorry, you said this already. But the quality composite index are the quality measures that you felt were related to the cardiovascular care.

(Ben Hamon): Right, they're the other HEDIS measures under the effectiveness of care domain. And we usually use the composite of those results for that same plan through their HEDIS mission to calculate the quality composite.

Nancy Garrett: OK. Do you have any (inaudible) weighting of these different categories to work in terms of the contribution to the overall cost? Is there any data that you have presented on that in the packet or can you talk through it a little bit?

(Ben Hamon): Well so, you know, again, we use standardized pricing so we don't have actual costs. And we do, you know, again, we don't really look at the contribution of each of these individual service categories necessarily to the overall. We do look at the correlations between these individual service categories and the difference components of quality and also in relation to each other.

And, you know, nothing has so HEDIS over the head, so to speak, over the last few years. But, you know, there's some very interesting and principally academic correlation that crop up and (inaudible) and trying to further understand that. But, you know, again, we don't make any assessments when these benchmarks are – when these results are presented whether plans, if they're over the mean for resource use or under the mean for resources in any specific service category as to that being appropriate or inappropriate because, really, again, this is just, you know, snapshots of their resource use for the whole year.

Obviously, their quality composite if they're under the average, we do judge because we do think that quality should be higher and that there's no excuses for that, really.

Nancy Garrett: So are the services category weighted equally or are they weighted according to the contribution to the overall resource use?

(Ben Hamon): As far as their contribution total medical, they are pretty much weighted equally because again, it's just – the results are all per member per month aggregate for each of the cohorts. Now it's basically just combined into a total medical. So there's no – there's no weighting for a contribution, it's just the data rolled up to a total medical.

Nancy Garrett: OK.

(Ben Hamon): So it's basically the total – a total PMPM for all of those different individual service components rolled up for each cohort that creates a total medical. We don't find any kind of weight because, you know, the cohorts are already risk adjusted. And I don't think each cohort has its own weight category applied under the (HHC) model. We don't do any additional weighting based on whether we think the, you know, the inpatient facility should count more towards, it's all the medical versus the procedure and surgeries and so on and so forth.

Nancy Garrett: Thank you. OK, thank you.

(Larry Becker): So this is (Larry). And so, at the risk of insulting somebody here. So OK, so these are a lot of numbers that you've pulled together, you created that chart, you showed us a minute ago with dispersion. Anybody gone back and sort of looked at the two ends of the dispersion in real terms and go create some kind of sense check of OK, so if I take these two organizations or these two data points and I compare them on some other measures, does this make sense that you can perceive a difference between these two organizations?

(Ben Hamon): So at the total medical level, I don't tend to hear about that. I do tend to hear about people's interest in specific components on, you know, certain health plans especially in places like inpatient services where they tend to be higher than the mean, tend to get very focused on why, you know, what their reasoning for that is and the justification for that is.

(Larry Becker): And so what do they tell you? Have you shown the people at the bottom or the people – well, you'd probably be more reactive to people at the bottom, you know, their resolve against sort of what the – what they're entitled in, what the top end would look like and say, "Gee, what is that and what's their reaction? I mean, do they say, “Ah”…

(Ben Hamon): You know, so the resources measures –well, I'm sure they'd probably say something little worse than the “Ah” in many cases. But the, you know, the resources measures are total medical expenditures for these patients for the entire year. So we don't only look for the ones relative to the condition.

So things we hear about are disease managing programs, population health programs, you know, investments they make in certain areas that might be driving their costs to identifying patients perhaps who might need services in specific areas. And then so that the, you know, we hear about the osteoarthritis some areas and some plans of a higher inpatient index for procedures and surgeries. And they show a very high quality on the same regards for some things.

And so, you know, they make those kinds of arguments to us. But again, like I've said, we don't really – we don't judge the thresholds, we just basically sort of show this as, you know, this is how you compare to your peers, it's very specific, and because, you know, these are individualize results, and calculate for your plan based on this data that you've submitted to us.

(Larry Becker): OK.

(Ben Hamon): So we hear a lot of reasoning why their results look the way they do, but, you know, so far they …

(Larry Becker): But if there's anybody said, "(Ah-huh), now I get it, now I know what I have to do different."

(Ben Hamon): Not that they told me, no.

(Larry Becker): OK.

(Caroline Perrier): This is (Caroline Perrier) again. I appreciate all the discussion that we've been having about the NCQA measure and I think that it's important for me to put it in a perspective that NCQA is measuring at the plan level. The other two measures that we're looking at are looking more closely at the provider level.

And so, again, I've heard NCQA answer the question about meaningful differences in terms of their output on their measure. But I still would like the opinion of others because I really have no frame of reference as to whether or not this is a meaningful difference and the one that I'm looking at specifically to comments that I'm looking at specifically are for 2431.

And under meaningful differences, it says based on the dry run of the measure in 2013, 7 percent of hospitals were below and 15 percent were above the national average. Is that generally a meaningful difference?

Female: Are you looking in some inputs from committee members or from us at Yale?

(Caroline Perrier): Well, from anybody that has an opinion because I just really don't have a frame of reference, so I'm – that's what I'm looking for. Thank you.

(Lina Walker): This is (Lina Walker). I'm a committee member and my question is very similar to yours, but I'm going to frame it slightly differently. And that hopefully, they can get us – our question one way or the other. When I look at the two measures, 2436 and 2431, what I don't see at the – I don't see whether variation is coming from and that makes it very difficult for me to understand how meaningful that variation is.

So, for instance, an example, suppose that this is a purely, there's absolutely no variation and also post acute (inaudible) and all the variation is coming from whether or not that one that – there is a readmission or to the hospital. So without, you know, without seeing – I mean, I don't know if a developer can provide more information about the sources of variation or whether or not that already is in there that somehow I'm missing it, because that would help me understand how meaningful those differences are.

Susannah Bernheim: Great, this is Susannah Bernheim from the Yale team, I'll try to answer both those questions and then you can let me know if I've address them.

So the first was just sort of how much variation is meaningful? And obviously there's no simple answer that we tend to use in our measures, pretty conservative approaches to identifying actual out wires, so as you note, there are not a huge number of hospitals that are out wires

But there is what we feel is a meaningful range in the distribution of the results for hospitals. So, I'm looking at our technical report which may or may not have the exact same years of data as you're looking at in the NQF applications. But, you know, the median hospital for the AMI payment measure, the full episode payments with standardize payments are about $20,000.

And when you go to the 90th or 10the percentile, you get a range of $4,000 difference. So, you know, when you think about – for instance, if you just brought all – if whole hospital found deficiencies that they all move towards the median, you would make a large difference in median outliers and you figure out how many AMI patients there are in this country with – for Medicare.

So, I think that a range between the 90th and 10th of $4,000 with a median hospital has an average of about $20,000, it's a pretty meaningful range. And yes, there are not going to be many measures, many hospitals identified as outliers here. But there's a pretty broad range of differences. And the second part of that question is where the difference is coming from.

And I don't think you have that information in front of you, so we can think about what we can bring to the in-person meeting that will help. But what we see is essentially it's coming both from variation in the index stay payments as well as variations in the post acute. So, I have something in front of me that has at least adjusted and so either unadjusted numbers.

But, you know, this – the median hospitals, the post discharge payments are about $2,300, but at the 96 percentile there, $11,000. So there's a very wide range in what happens in the post discharge setting. And there's also a relatively wide range of what happens in the index setting. We don't have

those risk adjusters, we don't calculate a separate risk adjusted index measure and a risk adjusted post discharge.

But we can certainly bring those distributions to the in-person meeting that helps people just to look at those unadjusted ranges.

Male: That would be helpful.

Female: I have a follow up question to that. So you wish you have – do you have spending on – in the post discharge setting and then you provided a really significant range. Now, is that all coming from post discharge setting, so are you including in that any hospital readmission cost, in patient cost as they were readmitted?

Susannah Bernheim: Right. So we include payments from any settings that the patient was in. So we look at readmissions additions to like hospitals, rehab hospitals, skilled nursing care, home house visit, outpatient visit, anything that's in the Medicare claims, except as we find out earlier the part D claims are included.

Female: I see. And then, when we – the information that you will bring to the in-person meeting, would you be able to provide kind of more assignments or more detail in how that payment is distributed across the different setting?

Susannah Bernheim: So I know that we have handy the – at a aggregate levels, sort of how much of the post discharge payments go to different care settings. So we certainly have that in aggregate level, I mean, we definitely have the distribution of the overall post discharge payments across hospitals. I think what you maybe asking for is that how much variation in there – is there, say, is how much payment goes to SNF across different hospitals versus readmission across different hospitals versus home health.

(Crosstalk)

Susannah Bernheim: Yes, I don't know if we have that or whether we can produce it in terms of the meeting, but I will talk to the team and see what we can do.

Male: OK. More specific because, you know, the part of the question here is where does the variation come from. And it's, you know, given that everybody is invasively the same DRGs, it's not coming from the hospital payment. But – so if you were able to break the group, you know, the population down into thirds of Quintiles and get the average spending and the variance in each of those cost categories in Quintiles so we could see whether, you know, high Quintiles were high across all or whether the numbers were being driven by SNF use and readmissions in the post acute …

Susannah Bernheim: Exactly.

Male: … period.

Susannah Bernheim: OK, that's a great suggestion. I can't promise what we can get done, because of the short time to in-person meeting, but we will try to get something close to that.

Just one thing to point out, there are actually the fair amount of variation even in the inpatient payments for a couple of reasons. Remember that DRGs are driven not just by the index cost, but complications of care which can lead to longer stays, if we carefully try not to risk adjust for so the hospital had higher levels of complication.

They may trigger higher DRG payments and also use of procedures, which an AMI can drive payments to some extent and if hospitals overall have significantly higher use of procedures even after risk adjustment that could also drive higher.

So we see a fair range in index payment as well. But I think you're asking a valuable question and we haven't looked at it exactly that way, so I'll see what we can put together.

Female: Thank you.

(Joe Stefanski): This is (Joe Stefanski) again. Going to back to measure 2158, the earlier Medicare spending for beneficiary, there is a data set that is made available to

the individual hospitals that shows by patient level or by – at the patient level spending in post discharge categories, isn't that correct?

Susannah Bernheim: Right, so we're not the developer of the measure, but that is my understanding, we similarly – when the AMI payment measure went into the dry run, every hospital could see a wide variety of information about each individual patient that was included and the post discharge settings in which they were seen over the 30-day period and what percentage of the payments will go into the setting or what …

(Joe Stefanski): And that is the sort of information with the patient, (what) the hospital really needs to be able to take action. And I'm hoping that you are going to preside the same kind of data.

Susannah Bernheim: Yes, we did exactly spent a lot of time trying to think about it as organize that data in the hopes that would be as useful possibly, you know, the way that teams are organized are not always intuitive, a lot of things get put into what's called the carrier files. And so we actually put a fair amount of time into organizing it so that it was more logical and more useful for the hospitals.

And so they will receive that and they did in the dry run of the AMI payment measure. And my presumption is that if CMS move these measures into reporting, there would be a very similar opportunity for hospitals.

(Joe Stefanski): I would like to suggest and this is a little bit to decide of looking directly at these measures, well, I think it's important. Right now, when data goes – it made available to hospitals, there's a readmission report which shows patient level. But it only shows some things about the original admission and then the readmission further down, but it doesn't tell you anything about what happens in between, where the spending for beneficiary type measures that we're looking at here start to tell you something about what services would provided in by whom in the interim.

And if those two kinds of data could be combined into a single data set for hospitals to use, it would be very useful. In other word, every line, every individual patient on the readmission report will have a corresponding record

in the spending report. And if they were combined together, it would just make things much easier for hospitals to be (aware).

Female: I don't know if CMS is on the line, but it's a good suggestion.

(Joe Stefanski): (Inaudible). Well, I'll bring that up again at the face to face meeting, because I think it's really critical, a lot of hospitals particularly the smaller ones don't necessarily have a lot of analytical talents sitting around to use. And CMS to be doing a great service for hospitals if they could make these data sets when they took them out on quality and that's easier to use. Or to decide correctly.

(Evan): Quickly, if more questions for developers.

(Lina Walker): Yes, this is (Lina), I do have another question. And I'm not a clinician so this is a question that will help me, you know, would provide context because I'm reviewing this measure.

So this is about the (inaudible) and the AMI measure. So my question is because I'm looking at 30-day episodes of care, from the beginning of admission, but I don't have a good sense of whether, you know, with (inaudible) and AMI whether there's a typical length of stay in the hospital or whether there's significant variation depending on complications that emerge, or whether those are kind of small rare instances.

So could you answer that set of questions?

Susannah Bernheim: Sure, that's, you know, the answer is both, you know, in general, we see in a Medicare claims data which is a particular population, the heart failure admission lasts between four or five days, but obviously, we see some that are shorter and some that are longer. And for AMI, it is a little bit shorter, I don't have that number off the top of my head. But there certainly are cases where the length of stay is going to be much longer because if you say, there can be complications or patients who are more sick.

But …

(Lina Walker): But that is so typical (inaudible) …

Susannah Bernheim: Sure.

(Lina Walker): Sorry. Hello. Hello.

Female: You're still on.

(Joe Stefanski): I think what we're dealing with is another one of those semantic things in the language used in the measure submission. I believe if we're only looking at the episode expanding through an entire inpatient stay plus 30 days post discharge, is that correct?

(Lina Walker): Well, (since) from the beginning of admission.

Susannah Bernheim: Right, so the admission day is the start of the 30-day episode.

Male: OK, that's …

Susannah Bernheim: And we had a lot of internal discussion about this because you can imagine doing it either way. But in general, what we don't want to do is have hospitals with patients who are sort of at risk for the outcome or this – in this case, you know, a risk for accruing cost that vary, you know, it's more fair to hospitals if we're examining patients for a standard length of time.

Now some argued that's generally the time the post discharge period, but then …

Male: OK.

Susannah Bernheim: … so anyway, the final decision that was approved by our TEP as well was that the fairest way to do this was to have a standard 30-day length of time and which the patient is being – that payments are being (inaudible).

(Lina Walker): OK. And you're saying that the significant like a modal number of day would be about four or five. And then there's some variation at (details).

Susannah Bernheim: Right. So I don't have those numbers in front of me, but – yes, in general, in Medicare beneficiaries four, five days is around where most heart failures admissions are and AMI are a little bit shorter. I think one of my team members is trying to find this actual data, if I get the exact number, I will get that back to you.

(Lina Walker): OK. And then, another context question for me is in your exclusion criteria. So, you exclude cases if heart failure or AMI is a secondary diagnosis. But you included if it's a timely diagnosis but it has some other secondary diagnosis. I'm just trying to understand number like at the proportion of cases (inaudible) they were secondary diagnosis. Is that somewhere in the – because I can't find that anywhere.

Susannah Bernheim: So I don't – I'm not sure what you're referring to. Our inclusion criteria designate a set of ICD-9 code that define patients that having come in primarily for AMI or primarily for heart failure. So we use the principle discharge diagnosis to identify AMI patients for the AMI measure and heart failure patients for the heart failure measure. We don't have a specific exclusion criteria about secondary diagnosis.

(Lina Walker): Well, if the patient has a primary discharge.

(Evan): Again, just as a reminder for everybody to please put their computer speakers on mute if you have your phone line open.

(Lina Walker): You do said that if a patient has discharge diagnosis of any other condition, isn't it – that this include the secondary diagnosis of heart failure or for the other measure AMI, so submission is not considered an index admission. So I assume you excluded those cases.

Susannah Bernheim: I think what that's intended to describe is that we don't count it as an AMI solely based of the AMI being on a secondary diagnosis code. We include patients in the cohort based on a principle discharge diagnosis. It wasn't meant to say, I can't imagine the situation if you have AMIs or principle discharge diagnosis and the secondary, but if that happen, you would not be excluded.

(Lina Walker): Yes, yes, yes. No, is it with primarily with something else, the secondary were AMI, you would exclude this.

Susannah Bernheim: I guess, you know, they – I think I'm confused by the word exclude. They would not meet our inclusion criteria, that's correct, we only include patient AMI measure if they have the AMI listed at the primary diagnosis. That's correct.

(Lina Walker): OK. I see. And I – so I guess this question is maybe for the clinicians and maybe even for – and also perhaps for the developer. Is there ever any uncertainty or discretion about whether the condition would be a secondary or primary diagnosis or is it pretty clear when it's going be coded as primary and when it's going to be coded at the secondary?

Susannah Bernheim: I can just say some of the sort of coding rule that the principle diagnose – discharge diagnosis is intended to be the reason that the patient initially was admitted to the hospital. Now whether there's discretion in that or patient come in with more than one potential discharge diagnosis. But that is how the rules are written and what's intended to happen.

(Lina Walker): I see. OK. Thank you.

(Joe Stefanski): And just to clarify that, I'll give you some examples of what not for CHF or AMI but examples of where someone came in for hip replacement and then had a heart occlusion and ended up going to bypass surgery. And the bypass surgery is the one going to drive to the DRG assignment in that case. And it wouldn't include that case in ortho, looking at ortho procedures. It will be excluded.

So if your AMI patient came in and saw while they were in there and broke a hip and they had some sort of orthopedic procedure, chances are they would be in an orthopedic and wouldn't be part of this measure.

(Lisa Lapps): Oh, I see. Is that correct? This is (Lisa). Thank you.

(Cheryl Danberg): This is (Cheryl Danberg). I had a question maybe more for the NQF staff as I'm trying to read through the documents. So, in the reliability section, it

seems like the emphasis for all of the measure developers has been around what I call reproducibility and less about being able to discriminate performance between providers. And I'm wondering how much is the expectation that there should be some of the latter included in this documentation.

(Evan): You know, we'll have to get back to you on that one. I'll talk to our methodologist on it and want to make sure that we're giving you the right information. I know we don't prescribe, you know, the specific method of reliability testing but that's something we'll take back and then we'll send out to the group.

(Cheryl Danberg): Right, because it seems like an – I think this was the case when we have the earlier meeting on resources measures that the focus is really on sort of the integrity of the data elements and ability to reproduce the results in these samples. But, it doesn't really tell us anything about whether there's enough variation between providers and what sample sizes are required to be able to discriminate performance to the extent that that its application.

Female: I think let's take from the Yale team that the insurance guidance is definitely that the reliability section is sort of test – retest as you've said, so that's what happens there. The place where we have our calibration and discrimination and lack of it is all in the section that's entitled – it's under the 2B4 section, if there's risk adjustment stratification and then 2B4.5 ask about sort of analysis to develop and validate the adequacy of the model and that's where we put those kinds of things.

We didn't do pure sample size but we did, you know, there's risk model discrimination and calibration and things like that. So you can find them there for our applications.

(Cheryl Danberg): Right. So I understand the calibration around the risk adjustment model but – because that, you know, reduces bias in the measurement but – and so that gets a validity issue. But, I think what I'm looking for is the extent to which you can truly differentiate providers on the basis of this measure.

Female: (Leslie), are you on the call? Do you want to speak to that?

(Leslie): Hi. Yes, I'm on the call. Well, I think – one thing I just want to be clear about is that, you know, our measure does not, in fact, compare amongst individual providers. We are actually creating a measure to, in some ways, compare a provider to do it on patients. So, I'm a little bit unclear or asking regarding how we might validate our comparison between providers.

(Cheryl Danberg): Well, I think isn't the ultimate application of this in the context of, let's say, a value based purchasing program where CMS or some other pair would try to bucket providers and, you know, to what extent would they be able to do that in a way that allows them to not misclassify providers into those buckets, if you will, those categories.

Female: So, right. And maybe what you're telling me is this is just an internal, you know, QI type measure, it's not for these higher stakes applications. But my senses is that, you know, the party is out there, want to use this in high stake applications. Is that right?

Female: Right. But I think, I mean – yes, I'm counting on (inaudible) but I think this is discrimination of the model, helps with that, right? And the ability to predict and range our results in the buckets so that they are placed in which we gave for the AMI standard measure, gives you some sense of those things.

(Leslie): Right. And we do calculate. It's in the NQF application, something that is called an ICC score and maybe this is getting at your question to test, you know, sort of the reliability whether we could calculate, in fact, a similar say risk standardized payment for the same provider using different samples of patients. So essentially trying to, you know, assure ourselves and everyone else about what we're getting at with the data that we're using in this model, the risk adjustment and the hierarchical modeling is that we are able to, you know, in fact, judge the same provider similarly across different data sets, which basically tells us that it's not particular patients in one data set that is driving how a provider may be "bucketed".

But, that – in methodology that we use reliably produces the same type of bucket for each provider over different data sets of patients. So does that sort of get at your question?

(Cheryl Danberg): Well, that's a very important measure. And I agree you want consistency and where you classify people, so some stability in that modeling.

But I think we're talking about different metrics. And I think that you've included some very important metrics. But I still think – and this is sort of my (Barger) question to NQF as it pertains to, you know, having measure developers to fill out these forms. You know, we think this interpretation of reliability is fold a bit loose in terms of what people are expected to provide.

Female: Can you say what you would be looking for?

Female: Yes. I actually can – (Evan), what I'll do is I'll send you this tutorial that one of the statisticians at RAND had generated around reliability. And think that if you want to circulate that to the larger group, that would be useful.

(Evan): OK, thanks. Yes, send it to me, I'll look at it.

Female: Sure.

(Evan): Do we have additional questions? Great, this time, we'll open up for public and member comments, operator?

Operator: All the lines are open.

(Evan): Great. Does anybody have any public or member comments?

Great, hearing none, we want to thank everybody for joining this call. So I especially thank the measure developers for taking time to answer question from the steering committee members and as well as the steering committee members and TEP members for evaluating the – reviewing the documents.

As far as next steps, again, a reminder that your preliminary evaluations are due on February 24th, so that is Monday, they are due by close of business. We will then have time to compile them and return them to you so you have time to review them in advance of our in-person meeting, which, again, will be March 4th and 5th here at the NQF offices in Washington D.C.

If you have any issues with the preliminary surveys, with your travel arrangement accommodations, anything relating to, you know, filling out evaluations or being at the meeting, please let us know and we will help you troubleshoot.

So, thanks again for being on the call. And this concludes today's call.

Operator: Thank you. This concludes today's conference call. You may now disconnect.

END