NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
February 7, 2014
1:00 p.m. ET

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

Katie Streeter: Hey, everyone. This is Katie Streeter our project manager here at NQF. Welcome to today's Endocrine Workgroup Number 2 Conference Call. Did everyone – let's just do a roll call if you could just say your names if you're on the line?

Patricia McDermott: Patty McDermott.

Katie Streeter: Hi, Patty.

(Bill Curry): (Bill Curry).

(Anne Watt): (Anne Watt) and (Cathy Gonzowski), Joint Commission.

(Tracy Breen): (Tracy Breen).

Ann Kearns: Ann Kearns.

Katie Streeter: Great. Welcome everyone. I'd also, before we begin, like to introduce our new senior director joining this project. She's not new here at NQF but she's new to the Endocrine Project. Karen Johnson, if you'd like to say hello?

Karen Johnson: Hi, everybody. This is Karen. I just want to (lay) any fears, Reva is still here with us at NQF but we just had to do a little shuffling around internally. So

now, I get to work with you guys on the Endocrine Project. So, looking forward to getting to know you guys and working with you on these measures.

Katie Streeter: Also just one quick reminder. This call is open to the public and we'd like to welcome (Anne) and (Cathy) who are representing Joint Commission who are the developers of the measures that we'll be reviewing today.

So what we'll do first is begin with Measure 2416. We do have primary discussion, (Tracy Breen) and secondary discussion of Patty McDermott. And before we walk through those measures, Karen will kind of give us an overview of how we see this call going as we review them.

Karen Johnson: OK. Thank you. As you know, we asked you guys to be prepared to be lead discussants for us. And we recognize that you might not know exactly what we are hoping for with that role. So for the workgroup call, I think I would like you as lead discussants to pretty much just do a real brief intro and description of the measure itself. And then, generally, as the lead discussant, we would have you actually go through the various criteria and talk about the issues that maybe you saw that came up or whatever.

And in our last workgroup call, I think like that was a little tough for folks that had never done that before. So at least for this first measure, I'll walk through the measure with you and kind of help be the lead discussant role. And that way, you guys will know a little bit more about what we are hoping you will be able to see for us in the in-person meeting.

And just so you know, this – we have a lot of time today for three measures. So that's kind of nice. And if you have never worked with NQF before, some of these criteria may be a little arcane or feel a little odd to you so we'll also use this time as a tutorial for a better word.

So if you have any questions about what, you know, the criteria mean or what we're expecting or what we are thinking, feel free to ask. It really is a conversation. And also, since Joint Commission folks are on the line as well you can certainly feel free to ask them anything that you want, you know, something that's unclear to you or you weren't sure whether it is something or

whatever. You know, just ask. This is your forum to ask the developer what they were thinking when they developed and when they submitted the measure.

So with that, let's go ahead and get started. So measure 2416.

(Tracy Breen): Excellent. This is (Tracy Breen). First of all, I want to thank you very much for managing expectations. I feel much better. I like being the guinea pig going first on this. So thank you. So, you know, without knowing some instruction about how the call is going to go, I thought I would just kind of review the very high level. In fact, what she said what kind of thoughts the measure and some of the things that popped out to me right away.

So this is looking at patients who have come in with a fracture who have laboratory tests ordered or performed at the facilities prior to discharge. And I think this is a really interesting measure for a couple different reasons but I'll come to it in a minute. So those specific lab tests are complete blood count, kidney function tests, a couple of things can give you credit for that, a liver function test, a couple of things can give you credit for that, serum calcium and then either the ordering of a 25 vitamin D or just the presumptive starting of oral vitamin D prior discharge gives you credit for that measure.

And right away, through where one of my first questions comes up because in the denominator so the total number of patients we're looking are patients the age of 50 or older who are discharged from the in-patient status with a fracture and the exclusion, age less than 50 years, comfort measures. That's fine. Enrollment in the clinical trial, laboratory testing performed in the threshold month. One of my questions right away in terms of what's not on the exclusion criteria that is on the next measure is whether or not people were already on therapy.

So – and this may be jumping around but on measure 2417 there are exclusions they actually exclude from the denominator, people coming on FDA approved pharmaceutical therapy for osteoporosis prior to the date of fracture. It just seems like that right away should be part of these denominator exclusions. So, you know, on the first page, that's what caught my eye but

kind of before we discuss that, one of the, I think, hang ups that people might have about this measure is the potential lack of evidence as to whether just doing this process measure because this is truly a process measure saying "Do you check this in patients who are coming with fracture who are at the age of 50?"

And the data – you know, some of – thinking about it, some of the false question because this laboratory data is really necessary prior to starting any therapy for osteoporosis. So if we're saying that one of our clinical concerns is that there's a gap between people who come in with a fracture, maybe their first, maybe then their second and that they're not started on therapy, and the ultimate goal of all of this is to get people appropriately assessed and started on therapy. You can't really start people on therapy in a safe and high quality way unless you have this laboratory data.

So I think that changes a little bit how we have to assess the evidence surrounding the testing of the process measure. I don't know if that makes sense to anyone or I also have to jump in on that process because I know, yes, there's data – one of the questions was is there data to support whether or not testing has been shown to improve the outcomes. And as far as I understand, there's data to say that when you look at all people coming in with fracture, there's a significant number of people with secondary causes of osteoporosis that would be picked up on this testing but there's no kind of prospective data to say, "This you test everybody (inaudible) you end up changing their course of therapy.

So I think I just like to (inaudible) the group is in very high levels of (offset).

(Bill Curry): So this is (Bill). I guess I would think of it in perspective of a patient who had been on treatment prior. Did they actually have the screening for secondary causes? And you would hope that would happen but I suspect that in fairly large number of patients taking these agents that perhaps it didn't or had they had recent testing to confirm that nothing has changed, have they developed chronic kidney disease, is there another abnormality causing their calcium metabolism to be off?

So I could understand why previous treatment or concurrent treatment might not be on the exclusion list.

(Tracy Breen): Good point.

(Bill Curry): But I don't have a huge concern about that. I think the point about will the process measure although it – I think it's a great way to screen for people that might have a secondary cause of osteoporosis. Will it impact the outcomes as you mentioned?

So it's a great process measure but looking at the process measure, will it impact outcomes? And I don't think there's evidence that would suggest that that – there's no evidence available for that.

(Tracy Breen): I would agree. It's a standalone intervention, just the mere process, the mere checking of labs is a standalone intervention. There's no data to support that it would change outcomes.

Karen Johnson. OK. So this is Karen. And you guys are already starting off really well. You're asking the right questions. So generally, how we'll do things in the workgroup and in the in-person meeting is we'll talk our way through the measure in the order of the criteria that we are using. So you've already started talking about the evidence so – and just to kind of get us all on the same page, this is a process measure and a criteria for judging the evidence of a process measure.

Generally, we are looking to see if there is some kind of a systemized review that gives information about the quantity, quality, and consistency of the evidence relating the process to desired health outcomes. So – and just a little bit of a – it's not a caveat but in some ways, you know, NQF would prefer outcome measures. But of course, there are many, many process measures that are very important to measure as well. So one of the things that we think about in terms of, you know, the different measures that are available is how close is the measure to the desired outcome with the idea that the things that are closer to the desired outcome and probably the things that you want to have as a national consensus standard.

So, that's why we start thinking about, you know, treatment. It's generally a little bit more (proximal) to the desired outcome or start backing up into things like assessments and things like that but that's further away from desired outcomes. And when that happens, it's not surprising that there's not a lot of evidence for those really distal processing. So with that, going through what this offers provided they talked about (inaudible) should lead to decreased morbidity and decreased readmissions. So that's their conceptual model of what this measure is supposed to do. Then, when you look at the evidence you should be – you would – what you would be looking for is evidence to show there's two things that there is some linkage between ordering these tests and decreased morbidity and decreased readmissions.

So they mostly worked with clinical practice guidelines. And they also included some other sources of evidence. So when we are going through and evaluating evidence we have the evidence algorithms that we have been given. And are you – you're bring that up? OK, all right. So I think for this as our kind of practice way of thinking things through, let's just walk through the algorithms and see where we land in terms of evidence and as other points come up that you want to discuss these other things, we can do that too.

So the first in our algorithm is, does the measure assess the smallest kind of health outcome. So what's the answer to that one? Sorry. This is going to seem a little elementary but procedure (inaudible) and another disclaimer, these are brand new algorithms for us. So we are all getting used to how these work. So I think going through might feel a little hideous that it might be a useful exercise for us.

Patricia McDermott: I think the answer is no. This is Patty McDermott, right? It's not – it's really looking at process. It's not directly looking at outcomes.

Karen Johnson: Correct. So that will take it to box three. So box three asks us for measures that assess performance on an intermediate clinical outcome process or structure so this one would fit, is it based on a systematic review and grading of the (inaudible) evidence or the specific focus of the evidence matches what's being measured. So that's kind of a little bit dense language but basically, there's a couple of things to pull off from that question.

One is we're really hoping to see – I want to say QQC, summary of the quality, quantity, and consistency of the evidence. And it's supposed to be the body of evidence. So, you know, you don't want to just cherry pick articles or something that may support your measure. You want to be sure that you're looking at everything. And then, finally, the other piece on there is does this evidence really match the focus of the measure. So in other words, are they – the reviews or the guidelines or whatever is presented to you, does it really relate to the measure for you, you know, the measure assessed by.

So if we go to the submission and I'm working my way there now. We'll go to the evidence form. The Joint Commission told us they laid out in a kind of a work diagram on how they expect the lab testing to lead to good outcomes. And then basically, they have a lot of things for you to look at in terms of evidence. So they have a clinical set of guidelines that has Preventive Services Task Force recommendation and then some other information.

So, if you look at the guideline's recommendations that they provided, they have one that R17, those are guidelines from AACE and R17 is the one that's really directly related to this measure.

Does that sounds right to you guys that it's R17 that is the salient one for this conversation?

Female: I agree so that includes the (inaudible) as well.

Karen Johnson: OK. All right. So they tell us that the recommendation has a level B and the evidence has a level two and then underneath they describe what B means and what the evidence level two mean.

So but over there's an answer – a question about a systematic review. So our first question is, is there a summary of the quantity, quality, and consistency? So for that, we have to actually go over to Section 1.7.

And since these are clinical practice guidelines, basically they're telling us that evidence level two again and there – it's R17s, it's the relevant one so we

know that there was one data analysis with a level two evidence and one study that's based on opinion (conceptually).

Are you seeing that? It's what I'm reading – what you guys are on there. Page 26.

OK. Now stop me if – I'm not going to go to this level of detail later but I just want to make sure everybody is comfortable with finding things in the forums and stuff. These are new forums to see. Now, there's new algorithms and new forum so everybody's getting these feeds.

So then the developers answer the question about quality in 1A7.6. There is no statement of that overall quality but it appears to be level two. So for quantity, it looks like there was (inaudible) quality and then consistency, they don't – this is unstated in guideline documents.

OK. So now let's go back to our algorithm and see if we can answer question three. So for measures that assess performance, is it based on a systematic review in grading of the body of empirical evidence? So how we can answer that one?

Anybody want to go (inaudible) there?

Female: It is graded, right? It's just the question is how strong is the value.

Karen Johnson: Yes. So these are based on clinical practice guidelines that has an evidence grade. So I think the answer there would be yes. And underneath there in the algorithm is, you know, it gives you ideas where it would be answered no, yes. So – and I don't think any of those apply.

So now we're in box 4 and it's a summary of the QQC provided in the submission.

So what do you think about that? Do you feel that they provided a good summary of QQC in the submission?

And again, there's …

Female: I'm trying to remember your comment. It's hard for us to see this because it's so small quite honestly.

Karen Johnson: Yes. I think we've got it as big as we can get it.

Female: (Inaudible).

Female: Yes. I'm actually – let me put in a plug and hopefully, you guys have had the chance to look at it, but I'm actually looking at page 38 of my committee guidebook where we keep the algorithms. If you happen to have that at your fingertips that would be worth to look at that.

Female: I'm going to go out and (inaudible) say yes. I think they have provided the summary of the data.

Karen Johnson: OK. And this is one of the things where it's, you know, it was a really question to answer and sometimes it's a little hard. So again, you know, we've asked experts to our table but as you guys have the expertise and, you know, but not everybody will end up saying the answer. So you just kind of have to decide for yourself if you think that's the case or not.

Does anybody disagree with that assessment?

Female: If we go back to what you were looking at just a few minutes ago where you actually said the citation. Did you say that it was – did you – was there an inference that it was a strong endorsement or not? I thought I heard that it was less strong, right? Again, it's hard to see what we're looking at here. We're going – you're leading us down a path of – you're leading us through the darkness a little bit here, you have to understand that.

Karen Johnson: Yes. So if you go to and (inaudible) I think is going to bring this up. If you go to page – hang on just a second. If you got to page 23 of the submission.

So what we asked them to do if they tell us the clinical process guideline, we would like to know and they does this for us. We've asked for the citation and then to actually give us the verbatim recommendation.

So they did that and so in the top corner of page 23, the recommendation 17 evaluate for secondary osteoporosis. And then the plan, they give you the grade which is grade B of the recommendation and then BEL and that stands for Best Evidence Level of two.

Female: So that's where you read evidence from at least one large well-defined clinical trial cohort of case controlled analytic study or meta-analysis? Next to is no conclusive level one publication.

So which do you take, the 2D1 or the 2E1?

Karen Johnson: Which page are you on?

Female: I'm on page 23. If you go down further.

Karen Johnson: OK. So D – right, right, right.

Female: So where it says D, it says D – right there. Above where your cursor is and then there's – it says D and then what's the inference of the 2D1 and the 2E1?

Karen Johnson: (Anne) or – and I'm sorry I forgot the other lady's name from – (Cathy). (Cathy), correct me if I'm wrong here but I'm assuming what this is, the description here is these are just the descriptions of how you get into grade B.

So what this is saying is that to be given a grade B, the evidence had to come from at least one large well-designed trial cohort or case controlled study or meta-analysis. Or there was no conclusive level one publication but there were at least one conclusive level two publication.

So in other words, either of these two things could have been true and that's how we landed in the B.

Female: That's correct.

Karen Johnson: OK. So that tells you how you got the grade for the recommendation.

So then for the evidence, you have to go to the level of evidence which is a level two, you have to go over to page 25.

Female: OK.

Karen Johnson: And it tells you what the level twos are. So if the meta-analysis is a nonrandomized perspective of case control trials, nonrandomized control trials perspective cohort studies or retrospective case control studies.

Female: Yes. I feel like we've answered box 4 and that they've, you know, (inaudible) the question is, is there a summary of the quantities of – is there a summary of the QQC of the body of evidence from a systematic review provided submission (inaudible). I feel like we've answered yes to that. The question is what's the quality of that work, right?

Karen Johnson: Right.

Female: So I feel like we're over at five at this point, right?

Karen Johnson: Yes.

Female: Whatever, OK.

Karen Johnson: Yes. So, pretty much depending on how you want to write these high, moderate, or low. If you know what the quantity and quality and consistency is which you think that you've gotten there. So you should be able to answer is the quantity moderate or high? Is the quality high? And is the consistency high? And if those are all true then it would be rated as high. I'm reading out of box 5A there.

So how do you know if quantity is high? Well, then you have to flip over. Hopefully, you have this up, turn it on table one, it's in the committee guidebook, that table one actually gives you the guidelines for how to raise the quantity, quality and consistency.

So for high, in terms of quantity, you need five or more studies. Do you guys have that table in front of you? OK. And committee members, are you guys seeing that table?

Female: Yes.

Female: Yes.

Karen Johnson: OK. You're telling me something is different. I'm not looking at their screen. Hang on just a second, let me walk across the room here. No. Table – are you on the steering committee guidebook? Go down a little bit lower. Hang on just a second.

Sorry. I keep looking at a different piece of documentation. So – OK. So now, you should see definition quantity, quality, consistency across the top? OK. All right.

So high for quantity is five or more studies. So – and for moderate, it is two to four studies. So we were told that this was based on two studies I believe, correct? On page 26 is the submission. Under R17, it was one meta-analysis and one, no evidence.

(Anne Watt): This is (Anne) from the Joint Committee and excuse for interrupting but we had a question too and that is, I realize that we're running through this, the algorithm for the guidelines, but then what about the other references, did they all get – I'm trying to learn this too. Do they all get rolled up into this analysis as well?

Karen Johnson: They do. Quite frankly, if you had kind of the order that it goes is if you have a systematic review of some kind and it tells you about quantity, quality and consistency, that's the first thing that we want to look for. And if you kind of have that then you don't really need to look any further. It's only if you don't have those things that you want to keep building.

Now, in your case, the guideline I think goes to 2003 (inaudible) in Section 1.8 if you had found additional articles that were post 2010 that were related in some way, you know, that's outside of the range of the systematic review that could come into play.

(Anne Watt): Thanks. Sorry to have interrupted and I know it's not our time but yours but that's very helpful.

Karen Johnson: Yes, you know, it is a little and again, these are new forms and they're new algorithms so we're all still learning where things are and hopefully, we'll all get better at reading them as we go. So this one is your call about whether you would call this moderate or low in terms of quantity because there is one meta-analysis listed there for R17 and then one study that said it is no evidence. So…

Male: I think it's because there's only one study that or one meta-analysis that's based on evidence. I don't know that I would say there's more than one study and the other is an – sounds like it's an opinion statement.

Karen Johnson: And you may recall that one of the things that we've said is, you know, we expect evidence to be empirical evidence, not opinion kinds of things. So, you're getting me. This is great. So the answer there is probably one or at least that would be how I might interpret that. If it's a one then going back to Table 2, and I know we're doing a lot of things here, one that it will give you a low rating for quantity. So, that fits into box 5B. So quantity can be low to high, so it doesn't really matter for box 5B, it doesn't really matter what the quantity is. This quality needs to be moderate and the consistency needs to be moderate or high. So our next question there is if the quality at least (inaudible).

(Tracy Breen): I would say, yes, still moderate on the quality. This is (Tracy).

Karen Johnson: OK, and do you want to – do you want to explain to your fellow committee members kind of how you landed there just so they understand what you're thinking?

(Tracy Breen): I think that there is, and again this – I'm having a hard time separating out the, the measure from what I know the study is because I know their studies very well and the measure. I'm still having a hard time of the measure being a standalone measure, just put out thinking about the overall care of these patients with osteoporotic fracture that the references that were submitted to support the measure were not necessarily studies designed just so we look at this process measure in a vacuum, right. I guess that's where I'm struggling a little bit.

We're using some data and some investigations that capture this laboratory data within them better part of a larger investigative question. I don't know if my confusion is shared with anyone, I'm just having a little bit of hard time just thinking if this process measure is a standalone from the overall care of this osteoporotic patient because that's really looking at studies we're looking at, right. We're looking at patients coming in with fractures, how many have secondary causes of – how many have secondary causes of osteoporosis, how many patients there are without care act for fracture but none of them are looking specifically at whether labs were measured, right.

So I know I'm asking a question that didn't answer your question but I'm still (inaudible) as I am trying to analyze the data for this particular measure.

Karen Johnson: Right. Well and I think that's a fair observation and I'll point you back to box 3 on that one. Because what we really need again is evidence where the specifics measures what's being – matches what's being measured. So, what we need is for these studies that you're familiar with they – in order to be able to speak to this measure and be evidence here for this measure, they need to actually speak to what the measure is about. So…

(Tracy Breen): And they all incorporate it as they go beyond that. I guess that's my question, right? Now every single, you know, thing they submitted to support it incorporates the measuring of these very basic laboratory data on the setting with fracture but that's probably the only thing they're thinking about is it's just a small piece so.

Karen Johnson: In some of the studies are they going forward, are they like evaluating a treatment or something like that? Something further along the care process trajectory?

(Tracy Breen): We talked about (inaudible) from the treatment study.

Karen Johnson: OK. So for that one, I mean, again, I think the evidence really needs to be specific to the focus. So, you know, having to incorporate assessment before you can get to the treatment, that's true but if the study in here are really about the treatment then I think that's what the evidentiary body is about, not about the ordering the test.

Female: So then, would that make in box 3 now other than yes, and you'd go down that lower path?

Karen Johnson: It could possibly, yes. Yes, if you feel like what's there doesn't really answer the measure focus then that is how that would work. So it is a little a tricky. I think …

Female: So then you'd end up with a low and a low?

(Anne Watt): Karen, this is (Anne). I'm sorry to interrupt again but could we tell you our thought process here since we're all learning?

Karen Johnson: I think that's OK this time, sure.

(Cathy Gonzowski): This is (Cathy Gonzowski) speaking. In section 8 which is the other evidence, I believe the strongest set of evidence that address this point about looking at the performance of lab test. The reference site has (inaudible) 2012 on the committee and medical association and that was a literature review of seven databases. There were 76 (RCTs), 24 meta-analysis and 34 other papers and it directly addresses the performance of laboratory tests.

There are one, two, three, four other references within section 8 that (inaudible) laboratory test performance and I'm wondering where in your algorithm you might be able to, you know, assess what is in section 8 and incorporate that in your determinations.

Karen Johnson: Just to clarify, (Cathy), and thank you for that. Can you tell us what page that is on, do you have that in front of you?

(Cathy Gonzowski): I think my page notes are different than yours.

(Tracy Breen): This is (Tracy), I'm looking at page 30 for anybody who, this is on the actual process measures, on the measure worksheet which I showed a bunch of (inaudible) and print it out but I'm looking at page 30, it's a version 6.5 (529-13) and the reference at the bottom is (inaudible) paper that you're talking about or the publication where it talks – I agree that it has a very large amount of data looking at the laboratory setting.

Karen Johnson: OK, so this is the second one from the bottom, the guidelines. OK, I'm just reading this for individuals with osteoporosis (inaudible) lab investigations warranted initially.

(Cathy Gonzowski): So this is for individuals with osteoporosis. Is that what this measure is about (inaudible)?

Female: Well, this measure is affecting people with fractures for osteoporosis, right.

Female: Right.

Karen Johnson: OK. So let's go back to our algorithms for a second and you have a guideline which are rated as a level with a B, I'm sorry I've already forgotten. Is a B and then evidence two. I think part of what's a little bit confusing is how to do the, how to identify the quality. And that's actually why we asked for a summary of these things. In terms of the various articles and things that were done after and (Cathy) released – provided a 1.8 because they postdate the guidelines review.

Female: I'm not sure …

Female: I didn't understand why you put what you put in 1.8.

Female: Oh, because some of the materials in 1.8 support the inclusion of Vitamin D and in prevention of fall which is contained in the evidence related to the United States preventive service effort evidence earlier within the submission form and others were chosen because they support requirements of laboratory test.

(Anne Watt): I think that – I think the broader answer, Karen, this is (Anne) speaking is that if it was under, you know, this is brand new process for us as well and it was our understanding that we should include additional information beyond the guidelines in that section and so that's why it's included in there.

Female: OK. No, I mean …

Female: I try to be – I feel like most of the data is actually in that section. Most of its accord is actually there. So I also didn't understand why it was separated out into some of the primary (and then supporting it) and then the additional of it.

Female: Right. Right. Well that's actually a very good feedback for us because you're right, you know, if you had – weren't bright systematic review that talks about ordering all these tests, it had to be done, right? You wouldn't really need to bring in pieces and parts from other places. And that's not always possible.

So I guess, you know, that then and I can even come close to pronouncing the author's last name, the one you said you thought was most applicable. If you use that information that was provided, can you make a determination in terms of (5A-C) about the quantity quality consistently.

Male: So, I have read this earlier in the week the concern that I had with answering that question was that, there are laboratory tests in this (inaudible) or if he or she says the name, article that are not listed in the Joint Commission's recommendation. So they talk about alkaline phosphatase. They talk about thyroid screening, they talk about continuous protein electrophoresis and yet these are not included in the process major as has been presented.

So, there's evidence that's presented but it doesn't match the measure as it prepared.

(Anne Watt): OK. And actually what just so we know and kind of get used to these criteria, what you're describing here is something that we specifically act under the validity section where we ask, you know, is the evidence, is the suffocation of the measure consistent with evidence. So if, you know, is it – kind of a different question and what we're asking at the beginning about evidence, but it is exactly what you're speaking to, you know, if there's evidence for these various different laboratory tests being ordered or being done, why aren't those in effect of the measure.

So, I'm not at – on, I'm sorry, I'm not sure that on the call that we're going to be able to decide considerably as a group where you should land in the evidence algorithm and again this – it is a judgment call to some extent, but, you basically, in order to use the (QPC), you have to know something about

the quantity, quality and consistency of the body of evidence and we don't expect you to go find all of the papers and figure it out for yourself.

So, you know, if you can't answer to your satisfaction but A, B, or C, then you can look at box six, you know, if you – just don't feel like you have the right information to answer box if that is A or B, box six just asked about the grade. Well, we know that there's grade and there the grade basically you need to have high quality evidence – sorry, I need to read this. High quality evidence, high consistency – this wording is a little confusing here.

Basically, you would answer no on that one, it's not graded high quality or strong recommendation. Probably it's his way to – basically what the first part of box six is it tells you what the (inaudible) (what brought you here) doing a – like the grade definitions which people can't use if they don't have to.

So another way to, again to find yourself if you don't – if you can't answer in box side A, B, or C, we'll get the grade and what would the grade tell you.

Female: (Anne), I just have a question about the process on this call, how would they use for the folks on the call, how would you like us to resolve this? Would you, you know, are we doing kind of forced (vote), high, moderate, or low or are we falling faster, what's – how do you typically run these?

(Anne Watt): Base calls are much less formal than what we will be doing in the in-person meeting. So in the in-person meeting, we will actually have this session similar to what we're having here about the evidence and what's present in the – what is presented or, you know, (inaudible) maybe concerning if there's any concerns and then the committee will actually vote on evidence. For the (one clip) calls we're not really nearly that formal and we're not going to come up at all with any vote asking to, you know, tell me what you think, you know, if you choose to, we can. But it's really just to kind of a group (inaudible) of, you know, where do you think this measure belongs on the evidence tree if you will.

So, we could do a real formal (inaudible) temperature taking if you want.

Female: Yes, thank you so my temperature is moderate on the data on this one.

(Anne Watt): OK.

Female: On this where I'm calling, for what that's worth, right? I'm feeling moderate.

(Anne Watt): OK.

Male: In my initial review, I thought it was moderate. I'm waffling between moderate and low, but my initial thought it was moderate.

Female: OK.

Patty McDermott: And I would agree with the statements, Patty McDermott. But I think there's – I wish I had a very truthful discussion and I think that there's an – our priority expectation that if we're doing something that we haven't pretty well nailed down, and I think we're concerned not as nailed down as I (get to be).

Female: There is a lot of room, you know, bringing in your own expertise as one of you said, you know these studies, you know. So, but other folks on, you know, many might know them as well, you are able to bring maybe, you know, your knowledge and background that other people might not be able to do. It makes more interesting committee discussion, you know, and different people will bring out different things.

So, if the committee pretty much agreed that it was a moderate, you know, in the full committee, then we will go on to the next criteria. If the committee pretty much that it's (below), then we would pretty much stop discussion at that point because evidence is a must have criteria. So that's how it would work in the full committee.

Female: Which brings me back to my original question about this whole process, right? So, let's say hypothetically speaking, right? So hypothetically speaking measure, whatever we are, what are we – 24.16, right? There's a standard of measure about assessing June laboratory task for evaluation secondary causes. So let's say whether it gets voted low, it dies in committee, right?

So, ironically then that brings us to the measure 2417 that I know we'll be talking about next versus that pre-treatment after fracture. Again my original

concern is based again starts therapy on some of osteoporosis without doing this lab test that we just said there's not in this hypothetical setting that we're not doing, right? So that's where I'm just having some trouble with the overall have the entire osteoporosis new measures come together, right? Because they're really not to be stand alone, they're kind of point, right?

(Anne Watt): Right. So, I guess let me just take you back to what NQF is trying to accomplish here. We are trying two endorse measures that are useful for both quality improvement and for various accountability application. Pretty much station Y that's kind of what we mean by national consensus (failure). So what we're not, yes, there's lots of really, really important things that need to be done in process of care. And as you said many of them have to be done in kind of a ladder-like structure and you have to this for us before you can go to the next thing, before you can go to the next thing.

So NQF endorsement decision whatever, you know, those things are still important to (pay). The question that we're asking you to help us decide on is, are those particular things important to measure and to actually, potentially be used – hopefully be used in both internal QI kind of applications as well as accountability application.

So, and that kind of take effects the proximal-distal thing. And it is – I'd better stop there and see if you at least – did that part make sense?

Female: It does, I should, you know, I'm always, you know, we talk about what's the risk of this, you know, here's my hypothetical situation here – here's my hypothetical scenario and I know we have our (J.C.) friends on the line (inaudible) helpful. If the NQF ends up endorsing let's say – I guess people coming with fracture, there's a recommendation that they are going to be either prescribes and counts or written a prescription for a FDA-approved agent for osteoporosis, right? And that becomes then a standard of measure by the Joint Commission for accreditation without the corresponding measure that certain laboratory processes need to put in to play. I could see the downside of that being there is a reflexive opportunity for hospitals to reflexively start therapies because that box get checked without that quality

piece that says you've also assessed them for secondary cause of osteoporosis, right?

So, (I've seen it) and the (inaudible) in the details, bear in mind. My favorite example is that it gets scores in the cardiovascular units for glucose, right? So the measure is, you know, within 24 hours of having open heart surgery, you don't want a single glucose greater than 200. That's great, but zero, right, is not greater than 200. So without actually defining really drilling down, I think there is potential for, you know, to walk down their own path, I guess as a clinician, that's my concern about this. Like I can't separate the two. I'm having a hard time separating the process from what I know is coming next in the therapy. That's – So I'm hammering on that point.

(Anne Watt): This is (Anne) from the Joint Commission and Karen may I make another comment just to sort of to frame this discussion a little bit more in terms of how the Joint Commission actually develops measures and stuff and they're intended to be used together and stuff for that very reason. So although these are all distinct measures, I ask you if I could do it and I'm doing it anyway and I'm sorry.

But – So, you know, we don't intend for any of these measures to be used, you know, to be picked. The reason why we develop measures in the steps is so that we can provide an overall for as close as we can come to an overall picture of the quality of care for particular disease entity. I don't know if that helps, but that is why you see these trees and they're all together and we use them all together.

Female: Thank you.

Female: So the challenge then going back to the task at hand is, if the first measure is (inaudible) low on evidence than and its measure set and the key goes beyond that point then, really this measure set needs to be going back somewhat to the drawing board at least to pretend better evidence before we go further.

(Anne Watt): This is (Anne) and again I'm going leap in here and say that although we use them as that, we ask for individual endorsement of the measure. The problem does not – if number one is endorsed and number two – I'm sorry if number

one is not endorsed, then number two or three are that's the Joint Commission's problem, it's how you are going to roll this out and if one of them is not endorsed in the other (2R), but we're not looking or asking you to review this we're asking you to review for endorsements the individual measure knowing that we use them together.

Female: Got you. Thank you.

Karen Johnson: And this is Karen and (Anne) is exactly right in that, you know, at NQF we look at measures for the evaluation purpose separately. So we want you to think about each one and think about, you know, is there enough evidence for the measure as specified to be called a national consensus standard. Now, you know, if offering these measures go down, you know, if that happens, you know, from the NQF perspective that doesn't tell the Joint Commission that they, you know, aren't allowed to use these measures. I mean they all lived, know what they need and what to do with them anyway.

But, you know, in terms of thinking about the portfolio of measures and, you know, you – there's many things that you have to balance and we're kind of getting a little bit off task here but you have to balance, you know, the idea of collection and burden and all these kinds of things as you're going through.

So with that, you know, let me just point you back hopefully and we will be getting a little later and more formally from you some feedbacks about what you thought about our staff review where we try to go through and pull out and point your attention to the really relevant piece as we think from the submission knowing that a lot of us here are not clinicians, I am not a doctor. (Reba) here is a doctor but she is a gynecologist I think, OB-GYN, not an endocrine specialist.

So, yes, there maybe things that we missed, but we try to make exact reviews pull out the relevant things to help you be able to make the evaluation. So we will be curious as to what you think about that. So that said, you know, just continue thinking about the questions for the committee. The evidence that's there isn't equally good for all laboratory test there provided in the measure. So again, you know, you need to have evidence for all of the things there in

the measure if there's only good evidence for one of the pieces for example that wouldn't quite see what we would need in terms of having an evidence-based for a national consensus standard.

Let's go on to the next one. And we are running a bit long so we'll try to go a little bit faster with (gap) and the reason that we're doing this again is that we all learn our new processes and as (Anne) has mentioned these are similar measures from Joint Commission so I think the next two will go a lot faster anyway. So perhaps in care, our self-review indicates that we didn't see any information about a (gap in care).

This when a (gap in care) can be a little bit tricky as a brand new measure but as it's a brand new measure and it's far to get data on that measure. And that's why we say data. If you can provide literature even if you don't have actual data from your measure, that would be appropriate as well.

So the question there for you is the (gapping care) and the rating that you give (gap) is actually in – if you have the generic scale. And what you'll find in this generic scale. OK. And that one doing right rate high, moderate, low or insufficient. So the question is, is there enough information for you use to use on (the data) and if so, then do you feel like there's an actual (gap in care). So how would you start thinking about (gap in care)?

Female: I think someone in the pre-workgroup comments I'm just looking kind of some with that – with a comment not assessed whether it's in general access, your first (inaudible) (fragility) fracture but not specific about lab testing meaning there's lots of data that say that in the setting of fracture, people don't take good care after, but there's no specific data around the lab testing (inaudible)

(Anne Watt): Karen, this is (Anne), what about the data that we present which regards the results of our pilot testing where we actually have these measures collected?

Karen Johnson: That data in the submission?

(Anne Watt): Yes.

Karen Johnson: OK.

(Anne Watt): I think we're looking for it right now.

Karen Johnson: OK. The short answer is we really need to see and attend to the gap. So let's see if we cab find it. So …

(Anne Watt): I'm sorry, I don't want to waste your time.

Karen Johnson: 6.9 percent (right off the call).

(Anne Watt): But we will have those data if they're not in the submission, we'll locate them and we will have that for the committee meeting.

Karen Johnson: OK. I do see on page 42 of the materials that we have and I realized you made me look in a little bit different page number, you did talk about in section 2B5 identification and specifically significant and meaningful difference in the performance, you have given some numbers and mean 16.6, standard deviation 20 percent, maximum 62.

So that would actually suffice for a (gap in care) set of data.

(Anne Watt): Or perhaps we just didn't put it in the right place.

Karen Johnson: Could be. Could be. And it was – we have to go back and kind of investigate and see how many – looks like looking real quickly, maybe six hospital, 130 records that's what it's based on.

OK. So, let's just see (Straumann) here, if we hadn't found that in the submission, then, you know, you would have no choice I think except to what's said insufficient information to be able to rate and we would have to stop discussion there in the full committee. Now that we see some numbers here, then we start asking ourselves, is that – is there a room for improvement? That's what (gap in care) are opportunity for improvement means.

So, you use your judgment there pretty much if there's a lot of variation or if there is this overall low performance you would probably feel although

everybody is different but you might feel with that is indication that there's opportunity to improvement.

Patricia McDermott: So, this is Patty McDermott, it's always the compliance rate with a mean of 16 percent? Oh, that's really quite low. And so with one of the things we just want to know is how that number was obtained, because I'm seeing that laboratory test performed in the prior 12 months for example is a criteria.

So if this is an abstracted grade, would we question how someone went after those – that lab testing, because if lab testing is done in an E.R. or an inpatient setting, you're not going to find it in the – let's go to each medical record to find that – let's find those results, else you have to go to those physicians which is that huge amount of abstracting if this is to be – if you're doing it as an administrative measure, we're just trying to capture the result – the evidence of testing and why is it done in a facility, you may not even find those claims part of an ER or an inpatient stay as it seemed done in the outpatient setting potentially, you can find those claims as long as the member – you know, I'm coming from the perspective of insurance plan where we have to imply – apply some kind of continuous enrollment in order to give the claim use – in order to know that the claims are available for us even touch, just some of the – just around this type of the metric and potentially how would this rate establish in your prior list.

Does anybody know or maybe I'm asking a question that doesn't come into this discussion but I kind of wonder…

Karen Johnson: So this is Karen and let me give a short answer and then see if (Anne) needs to add anything to this. This data is specified at the level of analysis of the facility so it's only looking at the hospital records but the data source could be paper or (EH4) data. That's how you specify the measure. Correct?

(Anne Watt): Yes, that's correct. And each individual chart is abstracted and the way that we derive the 16 percent is we run a pilot test for I think six months where volunteer hospitals adopted this measure set and actually collected the data on their osteoporosis patient. And the numerator is – and the denominators were

calculated according to our measure calculation algorithm which is in the submission. And that is the rate that was derived.

Female: So if this testing was done in the doctor's office in the prior 12 months and the only other question I would bring in is if this patient has had osteoporosis for a long time, that time of testing may have been done five years ago to determine whether there was a secondary cause. The hospitals would not get credit for the knowledge that that secondary testing had been done unless in some way they're asking that question each time a patient with osteoporosis is admitted.

(Anne Watt): I'm going to answer. This is (Ann). I'm going to ask (Cathy) to explain the exclusion. Those patients would not be included in this measure because they are specifically excluded according to data elements.

Female: The patients who had had laboratory testing that could be found either in the electronic record or, (Cathy), in the paper record. If they had had all of these lab tests in the prior year, they were excluded from the measure. So they were neither counted against the hospital nor for the hospital.

If the lab test has not been done in the previous year, it would be essential to repeat. If the patient is in, for example, a multi-facility health system, usually that information was available. It was an integrated medical record.

Male: So I guess I'm confused. If the patient fell in Missouri and their medical records were in Pennsylvania and they had the review for the laboratory work and it wasn't done in Missouri and they couldn't find it in Pennsylvania, they would still be in the denominator data, correct?

Female: Yes.

Male: All right. And then in the table that you presented, table 3, and the data element validity, if I'm interpreting this correctly, then it's about 60 percent of the time that laboratory data within the last 12 months was able to be found by the abstractors, is that my – correctly …

(Anne Watt): This is (Anne), and I know, we're looking for it but no, I don't believe you are.

Female: No.

(Anne Watt): If we're talking about reliability that says that the – when we do reliability testing we send staff out to actually re-abstract data that was already abstracted by the hospital and compare them. So that's what we're talking about in the reliability section. And it really has nothing to do with how frequently it's there. In the validity table 3, what you're seeing there is an assessment by those who were actually testing the measures as to whether the information was clear or collectible and if we were asking for the correct data source. And so, for the specified laboratory tests you can see ratings in the 90s, lab test order to perform before discharge bigger than 90 percent concurrent with our approach.

Karen Johnson: So I think going back to the question about the 60 percent that is in that table, that would be something that you would consider under validity. So these (sector) saying that, if I read this right, it seems like it might not be that collectible. So then that raises the question of the validity of the measure and being able to actually consistently – well, not consistently, but, you know, are the results going to be what you think that they are.

They really reflect quality in that is not collectible. So that's – the validity (partner), even in the evidence to think about, it has – there's all these different threats to validity you have to think of. But that is a good question you definitely want to be thinking about.

Now, I really hate to do this but we're going to have to really push a little faster. But I think this is still instructive. Hopefully it's not too tedious for everybody on the phone to go through slowly. Let's go the – let's skip ahead to the scientific acceptability.

So we ask you to think about the specifications specifically and then to look at reliability testing. So the reliability testing looks like, and (Cathy) had already said, the inter-rater reliability, they have to two different abstractors for the data and took care of what they found. And they're testing. It was predicted on a 133 patient charts (inaudible) six hospitals. And they report a high degree of agreement. So I'm assuming that is a percent agreement statistic.

And then they provide one Kappa score. So just to remind of you our data – our reliability testing requirements, basically we allow the developers to test that, either the data element level or the score level or both. So for this particular one, they did it at the score site at the data element level and they inter-rater reliability testing, which is definitely appropriate.

And the other thing, we were doing data element testing it it's important that you look at the all the critical data element. So one of the things you have to think about is all the critical data elements testing. And basically that means they're all the things that make up the measure were those looked at.

And so you look at the critical data elements, then you think about the test sample. So they said that they looked at a 133 charts in those six hospitals, and they gave some information about, you know, the distribution of the hospital, so that sort of thing. So you consider whether (inaudible) that is a reasonable enough sample size so that you feel comfortable that the testing that was done, you know, we won't use the term (generalizable), you know, we're not talking about statistical (generalizable). But, you know, is that a reasonable enough sample to feel good about the reliability?

And then once you make your way through that question then you consider the results of the testing.

So with that let's take a look at the reliability, or let's just walk quickly through that one.

So (inaudible) that up, Katie? OK.

And does anybody have any questions while we're bringing that algorithm up? OK.

Box 1 has to do with the specifications. And for time we won't go into the specs here. But you have to think, you know, in the in-person meeting, if there is something that you're concerned about with the specifications, you know, similar to that, questions about the denominator that was asked early on, that might be a point of discussion.

So let's assume that the specs are great. And so that puts us in Box 2 with empirical reliability testing conducted using tests with the measure as specified. And in this case I'll go ahead and answer that for you. I think that it was – sometimes people get confused about what reliability is and they'll give you the performance rate, you know? They'll give you that 16.6 percent, or maybe they'll give you the 16.6 percent in one year and, you know, 17.9 percent in the next year and that sort of thing. That's not what we mean by reliability.

But the testing methodology that's ultimately used is (inaudible).

So then that takes us to Box 4 with reliability testing conducting with the computed performance measure score. So the answer to that one is no, they did data element testing. So takes us to Box 8.

And so then the question is was it conducted with patient data elements that are use to construct the measure score? So that's where you have to think about, you know, with the critical data elements that, you know, are critical for calculating the measure looked at.

And so we'd have to go through and we don't have the time now to go through each one. But that's something that you'll do a little bit later. You know, just look at all the things that need to be looked at and they looked at some.

And then question 9 "Was the (med) inappropriate?" So they did do inter-rater reliability which we said is an appropriate methodology. And generally, a lot of folks will give perceptively with statistics, and those are fine. They give you some information but they don't give you as much information as you might need because really the percent agreement has – can be really high just by chance alone. So that's why you – we'd expect to see some different kinds of statistics. And often the Kappa statistic is one that folks will use.

So they do give one Kappa score. I think the thing that's real confusing for me, I think, in here (Cathy), to help us with this, there's only one Kappa score, but really there should be a Kappa score for each of the data elements. So …

(Anne Watt): Sorry. This is (Anne). Maybe that's just a function of the way that we reported it. We do, as indicated, every data element is assessed for agreement rates. And we just roll up the Kappa score to the measure score because obviously the critical measure – I'm sorry, the clinical data elements that are required to compute the measure are there, and so that's why we do it. Because there are so many data elements that are superfluous.

So that's just why we did that.

Karen Johnson: Is that one rolled up score like an average Kappa score? Is that what you're …

(Anne Watt): That's the Kappa score for the measure rates themselves when looking at all of the Kappas of all of the data elements.

Karen Johnson: OK. So I'll be honest in that I really – I can't visualize a rolled up Kappa at a score level. I don't quite understand that one. But let's just move on to …

(Anne Watt): How about if we have our biostatistician call in for the committee meeting? So if there is that question he can answer it a lot better than I can.

Karen Johnson: That might be worth it. Or, you know, maybe just send us an e-mail or something like that. I might be the only one who's confused about that. But that would take us, let's assume the answer is yes there, that would take us to Box 10. And then you have to consider the reliability statistic and the scope of testing.

So for the sake of argument we would go with the 0.728, and you have to decide, committee members, whether you feel that that is either – is that rate high – or not high, moderate? Is it high, the rate, within not only data element testing. And or you feel that it's low?

And when you're thinking about that you think about the results as well as the testing that was done. Again, that was at six hospitals and 133 reference. So do feel like that's, again, an adequate sample to be able to make some determination about the reliability question.

Male: So the thought that I had was there were, as I recall, like 24 hospitals that started off on the journey, and about half of them continued on the journey, and now we've gotten just down to a half of that. So a quarter of the total number of institutions or enterprises that were an initial invite with the final testing of 133 patients. So was there bias that has been created there that these are hospitals that were the most motivated to make adjustments and changes? And then try to get to the data where the pilot testing is for or the Joint Commissions?

(Anne Watt): This is (Anne). Could we just clarify something, please? There were records of more than 2,000 patients in this test. The 133 is the number of records on which we did our integrated reliability in that we went out and so we looked at the records. So these rates represent the rates in 2,000 charts not just the 133.

Male: But still up to six hospitals?

(Anne Watt): That was the sample of these patients in those – yes, we visited six hospitals.

Male: OK. Thank you.

Karen Johnson: So, (Anne) (inaudible), to make sure I understand because I don't think that sinks through to us either. You're saying that you had a lot more hospitals so we're hoping you're testing but you actually only did testing at six of the hospitals in the 133 records.

(Anne Watt): Yes. We only did inter-rater reliability testing. All of the hospitals were tested – or tested, queried in terms of the availability of the data with ease of data collection, the general validity type issue for the measure.

Karen Johnson: OK. So the tables that you have in the validity section actually represents a lot more hospitals than just six? And …

(Anne Watt): Yes.

Karen Johnson: OK. OK. So you have to think about, again, the value there and how many records were looked at, and make sure you (inaudible) as to whether you would rate it as moderate or low.

Are there any other questions about the reliability?

Again, pretty straightforward, except for me. I think the question would be that rolled up Kappa. Generally, Kappa is used for a nominal kind of yes or no kind of thing. So, again, rolling it up – let me put it this way.

Seeing the Kappa scores for each individual data element is what we would expect to see.

(Anne Watt): You know, excuse me, this is (Anne). And I don't think that the argument (inaudible) but actually we discussed this methodology with (Karen Pace) in a recent project. And this seemed to be what she was looking for. And so I'm not entirely certain that I would say that that's not an appropriate thing to do. I'm just obviously not explaining it very well.

Karen Johnson: Maybe the inter-rater reliability methodology is definitely appropriate, and calculating Kappa scores is definitely appropriate. My question is, I just don't – I don't really know what that overall Kappa score is because you can also do the Kappa scores with the individual data elements.

(Anne Watt): And we'll have the statistician explain it, Karen Johnson:. I'm sorry. I just – it's not my area of expertise. But we will have the answer for you.

Karen Johnson: OK. So committee members, are you ready to try validity now?

Male: Yes.

Karen Johnson: OK. So validity testing looks like – you said safe validity for the data elements. So safe validity is what you did.

So let's look at our validity algorithm in a minute, Katie, or (inaudible). Thanks. Hold on just a second.

So, again, testing can be done at the data element level at (inaudible) measure school level, or for validity we do, well, safe validity testing. So we'll short cut through the algorithm here.

So Box 3 was empirical testing done. The answer from what was submitted is no. Then that takes you to Box 4 with safe validity (systematically) recognized expert on the computed performance measure score. So the directions in the algorithm – and what we mean by that is, you know, it's the score that it produced, the safe validity that we're interested in is can the measure be use to distinguish good versus poor quality.

And then answer no if it's focused on data element accuracy, availability, feasibility or other topics.

So the fact review question there is it's the same question as the sample size which was already addressed to some extent. And the results demonstrate sufficient validity.

So you have to ask – and you have to figure out if you answer yes and go to 5 and think about the results, or you answer no because the safe validity was done not on the score itself, but actually on the data element.

Female: I think that was the meat of the conversation, wasn't it? That it was done on the data elements? Or I think, I'm just understanding (inaudible) the conversation has been?

Karen Johnson: That's how it appears in the submission.

Female: Right.

Karen Johnson: OK. Let's go ahead and move on quickly. Are you guys feeling comfortable at least with the process and feeling comfortable, at least more so with the algorithm than perhaps you were when we first (inaudible)?

Female: Just take a step back. The idea here is that we're doing a preliminary review of what we've seen in the submission. This would be, at the larger meeting, will be presented to everyone. (inaudible) even not the only people voting on these concepts. It would be the entire committee voting on the concepts, correct?

Karen Johnson: Correct.

Female: We're just doing a preliminary assessment?

Karen Johnson: Right.

Female: Right. Thank you.

Karen Johnson: Yes. Now, we have spent about an hour and a half working for a way through this one submission. In the meeting we'll probably have 20 minutes or so – to get through the entire thing. And that's another reason that we want to (inaudible).

Female: Through all three?

Karen Johnson: No. Generally about 15 to 20 minutes is what we allow our self. Sometimes a little bit less.

Female: So then the question will be, as we've gone through this first one, are the statements in the approach to the things we just discussed similar in the other two measures?

Karen Johnson: Right.

Female: Right.

Karen Johnson: If they are, and oftentimes, I mean, all three of these measures are Joint Commission measures, I think they're almost identical in terms of how they tested in the different things they were testing. So, you know, once you kind of decide to how you're going to do it on the first one then it's kind of the same – or some of the things at least is the same, or you have the same answer you go a lot faster.

Yes.

Female: So that's why this pre-discussion is certainly (inaudible).

Karen Johnson: Yes. Yes. It's hopefully doing a couple of different things, you know, just getting everybody familiar with the measure. You know, doing a deep dive in

the measure, but also getting, you know, more familiarity with our process and the criteria themselves.

And as I said, you know, working our way through these algorithms and that sort of thing.

Just going to the next – I'll be a little bit brief with the next criteria. One is feasibility. And feasibility – feasibility and usability are not what we consider a must pass criteria. So there's a lot more, even more room for you to bring your own value in terms of how feasible things are and useable.

So, again, with the questions that you'd want to think about are the data elements routinely (inter-rated)? Are they available in electronic form? And is the data collection strategy ready to be put in to operational use?

Now, some folks may kind of think these paper kinds of record measure because they're not as (inaudible). It's more work to do, paper medical record measurements. But, while that may be true, that's certainly, you know, what would be a reason to, on its own, to bring a measure bound since there are very many, you know, great paper-based measures.

And so, again, feasibility, you have some (wiggle room) in there. Looks like the workgroup comment had to do with the Vitamin D level taking some time. That might be a question for (Anne). Do you have a flavor of when these paper or charts are looked at?

(Anne Watt): When are the charts looked at? They're looked at when the patients have been discharged for 30 days. And it used to be that (25) reviews were done mainly in reference laboratories. Certainly that was the case when we alpha tested these a number of years ago.

However, in the interim, there are a number of facilities who have developed the ability to do in house testing rather than at a reference lab or a university lab. And so the turnaround time in some facilities is the next day, sometimes two or three days after the drug.

Karen Johnson: Another thing that we talked about – and, well, actually some usability and use (inaudible). Let's go on the usability and use very quickly.

This one's a little easier to answer, (Anne) and (Cathy), because it's a brand new measure. So it's not currently in use we don't ask the questions about improvement necessarily. But it's not in use so it's not publicly reported. It's not yet used in one accountability application.

And these questions are there because our criteria for usability and use is when (met) after three years of being endorsed, we look at them to see if folks are actually using them. And, you know, if they're not being used in, you know, these kind of accountability programs after a certain amount of time it tells them the question whether or not, you know, we need to continue endorsement for them.

So those are kind of move point questions for brand new measures that are not yet in use. But what is a valid question for even the new measures is, will the benefit of the measure outweigh any potential unintended consequences?

So I don't know if anybody has anything they want to address on that. Did anything kind of jumped out at you?

OK.

Let's go ahead and stop talking about this measure and go ahead to the next one. And let's just walk through. We want to have our (lead discusser) tell us what this next measure is about. The (312-17).

Let's see. Are you ready to do that, (Anne)? Or do you need me to do it?

(Anne Watt): I'm sorry. Were you looking for us to – this is (Anne) from the Joint Commission. Were you looking for us to introduce the measure?

Karen Johnson: Yes. Actually I was looking at the other (Anne Watt):

(Anne Watt): Sorry. Sorry.

Karen Johnson: She may have had to drop off. OK. Let's just go quickly through this measure. It is patients aged 50 or over with a fragility fracture who have either a DEXA scan ordered or performed or prescription for pharmacotherapy for osteoporosis, or who are seen by or linked to a fracture liaison service prior to discharge from in-patient (desk).

If DEXA is not available then you can – another specified fracture assessment method may be ordered or performed.

So thinking about the evidence, let's just kind of talk through a little bit quickly. Is the evidence equally strong for ordering a DEXA scan or providing medication or referral to a fracture liaison service? And are those care practices closely related to preventing future fractures?

(Anne Watt): Basically what it's saying is are they being treated or continue – has treatment been initiated or has just been done to determine if they truly have osteoporosis or not. So, you know, I think you're covering the scenario of checking for and treating …

(Anne Watt): Hello?

(Anne Watt): (Inaudible).

(Anne Watt): Hello?

(Anne Watt): Hello?

Karen Johnson: Is that you (Anne)?

(Anne Watt): Yes. Hi. You couldn't hear me.

Karen Johnson: I couldn't hear you. We thought you had left us.

(Anne Watt): No, no. I was waiting. I'm sorry. So I missed what's already been said because I just re-dialed in.

Karen Johnson: OK.

(Anne Watt): So, do you want me to start from the top and be brief?

Karen Johnson: We already did a real quick introduction of what the measure is. But do you want – so you've kind of the general (inaudible) of maybe the points about the evidence? Are you comfortable doing that?

(Anne Watt): Sure. Sure. I think that the evidence that treating people with low bone mass (presents) fractures is pretty clear. I think that part of this measure – I think of the point of the measure or the intent is good. I think that – what I found a little confusing is doing a DEXA scan is not the same as initiating a treatment.

And so I had a hard time understanding. Some of these seemed very close to the outcome and some were a little bit distant. So I think a DEXA scan is close than lab test but it's still not treatment. So I think that's a different element. And there's even less evidence about treating people with a normal DEXA scan without a fracture.

So I think the evidence for treating people with osteoporosis without a fracture is high. And I think that there's multiple studies that address that type of evidence.

Karen Johnson: OK. So what you need to do is – and we don't have time this time to go through each of the things that the commission has given in terms of the different guidelines and papers and stuff. The question, you know, is mostly evidence and a strong evidence is about the treatment. Is there similar evidence for ordering the scan or referring to a liaison service? Because you are evaluating the evidence for the measure as specified.

So, you know, you need to have evidence for all the pieces of the measure.

(Anne Watt): Yes. And I think I can go through that, but I know our time is limiter. But I think there is very good evidence for the liaison service, and, you know, whether they have all that in there, and the easiest way to find I think is what I found for all three of these measures was some of the data was in kind of a separate section that I found later in my process.

So as we talk about (inaudible) first measure.

Karen Johnson: OK. So you feel like it might have been more toward the end of the evidence review?

(Anne Watt): Yes. Starting with the practice guidelines kind of was a little bit off track with some elements.

Karen Johnson: OK. OK. All right. Any other discussion on evidence for this measure?

No disagreements, no questions, no – OK. The gap in care.

(Anne Watt): I guess I think I'd say the rest is pretty well both with their pilot and the body of published works and that it's not well addressed at the present.

Karen Johnson: OK. And so you're harkening more to their testing results that they – so those are – I'm just trying to be clear here.

(Anne Watt): So there are published studies saying that people with fractures don't get assessed. So I think it's the combination of both of those that …

Karen Johnson: OK.

(Anne Watt): I think there's a clear gap with the measurement, their pilot or the published data.

(Karen Johnson): OK, all right. Reliability. This one is like the others, the hospital level and data elements.

(Anne Watt): Yes, I think that I have a harder time separating the reliability in terms of can you find, you know, are they clearly defined? Is it logical? Is it going to be easy to meet measures? DEXA is not always recorded in a hospital level chart.

Karen Johnson: OK. Where is that found usually?

(Anne Watt): Well it's often done in a physician's office or a private clinic, or if not – in my institution you don't do it as an inpatient.

Female: Right.

(Anne Watt): So ...

(Tracy Breen): This is (Tracy). I totally agree. It's not an inpatient …

(Anne Watt): No.

(Tracy Breen): Inpatient (inaudible) would ever get. Actually, no.

(Anne Watt): You can't send out. The machines aren't at the hospital in my institution. So you can't (inaudible).

(Tracy Breen): And you would never delay someone's care to do that part of (dispersion).

(Anne Watt): Right. They allow for just the order to do the task, but I think when you are abstracting the data and saying have they had one in the last 12 months, it gets very (inaudible) as it's not going to be in the hospital record easily, unless it's a very well connected electronic network.

Karen Johnson: OK. So if we were looking at the reliability algorithm, we kind of breeze through that section in the last one, but the very first box asked are submitted specifications precise and ambiguous and complete so that they can be consistently implemented.

So I think your question about whether you would know about a DEXA scan being used, that would be something that you would consider definitely under that section. And then you would answer the question in Box 1 based on how you feel about that. Is that a killer for the measure?

(Anne Watt): Yes, I don't think so. But the other question I had and maybe the secondary reviewer could – I know we're backing up a little bit, is there was this statement that other measures other than DEXA would be allowed, and I could not figure out what those are. So they could have some other risk assessment, and that I couldn't figure out what those were.

Patricia McDermott: Guys, this is Patty McDermott. I think the ability to find that DEXA scan is critical to the integrity of this measure, and whether there's a big – the idea that it's ordered is often hard to find, I would suggest in a medical record.

And but the fact is it's done two weeks later or, you know, a month later or if this was done in the prior six months. I would say that you already know the status of the members or the person's phone, and which you need to repeat one just because they fall in again.

So I think those kinds of things will generate conversation.

(Anne Watt): I agree. But I think they allow for you to just treat. So if you found the DEXA that was done six months ago and they weren't on treatment, you could just initiate. You know, do the proper evaluation and initiate treatment, and that would meet the measure.

Patricia McDermott: Also, if the person with the DEXA was negative then they're not going to be treated. But they have done what they were supposed to do because it's really not after his visit.

Female: Right. True.

Patricia McDermott: I'm just trying to be a little broad in the – think about things from, you know, what can you really collect and how are you potentially going to give a provider a negative hit and you're going to come back and say, this doesn't make sense. Do you want me to do another test that cost X number of dollars when I just did one?

(Crosstalk)

Male: Are we talking about reliability or are we talking about feasibility?

(Anne Watt): Yes, I'm not sure ...

Patricia McDermott: What if they're both?

Male: Well, I think if you look at the specifications, I think that they're fairly precise and ambiguous and complete. And I think that you would have to answer yes to that. Whether it's feasible to collect that data I think is a different question.

Patricia McDermott: Good point.

Karen Johnson: I think that it is – you have to think about that question actually, under all three, when you are thinking about the specs, the validity and the feasibility, so it's all three. I guess, really, what you're looking for is being able to have consistent results across hospitals.

So, you know, it's not – for this particular question that you've brought up, the question is more not the precision of the code but, you know, can they be consistently applied? And, you know, I have no idea. That's why you guys are here.

(Anne Watt): Right. And I think the answer to that is it's difficult.

Karen Johnson: OK. So yes, you can see – and it's actually, it would be lovely if we can talk about, you know, issues only within our little role of, you know, (inaudible) as you pointed out is usability, and it's also validity. You know, how valid is the score for (inaudible) hospital if they can't, you know, can't do things consistently or whatever.

So, again, we're not really coming to a kind of workgroup conclusion necessarily but we are kind of starting to air the questions and concerns that you may have. And it also really shows the need for you guys because when we give the questions to the committee, I didn't know that DEXA scan results may not be easily retuned. That's not something I would have thought to ask about.

Was there any other, I won't say burning questions, but any other concerns or any other questions that you would ask (Anne) and (Cathy) while you have them on the phone about this measure?

(Anne Watt): I would just like add from what the other risk assessment things are because when I looked at the supporting data, I couldn't – are those codes or numbers for quantitative ultrasound or QCT? I don't know what those other measures, risk assessment measures that would be acceptable are.

Yes, did anyone else find that information or am I the only who thought it was not clear?

(Cathy Gonzowski): (Cathy). Those methods are found in the appendix in Table 6.1. And they are DEXA of spine, or of the proximal femur, a QCT of the spine, QUS of the heel, DEXA of the fore arm, a DEXA of the heel and the fracture. And we are working on refining the ICD-10 codes that are supplied with those.

(Anne Watt): But when I looked at table, it was hard for me see that those were what was listed. I thought they were a bit. Thank you for clarifying that.

Karen Johnson: And, (Cathy), so I understand. I'm looking at the appendix of the measure submission. Are you talking about a different appendix?

(Cathy Gonzowski): No. It is appendix Table 6.1. I didn't see it easily there. So, maybe the version we have is not ...

Male: I couldn't find it either.

(Cathy Gonzowski): Yes, it's just a list of things that I never heard of.

OK. We might not have the right answer.

Female: But again.

Female: OK.

Karen Johnson: All right. There's something else. Any other questions about this measure?

All right. Well, let's go on to the third one. 24, 18, discharge instructions emergency department. And I think, Dr. (Curry), is that yours?

(Bill Curry): Yes, so this is looking at patients who have been seen in the emergency department with fractures. The expanded fracture list include vertebrae, pelvis, wrist, humerus or ankle. And that they have gotten either discharge instructions to the patient or caregiver stating that there's a need for follow up with primary care physician or other specialist physicians, or outpatient hospital department for possible osteoporosis. So there's the risk of future fracture, or they have been seen contacted by or linked to a fracture liaison service.

So it's pretty much the same exclusion, less than 50, comfort measures only, participation in the clinical trial.

So, I think that the Joint Commission has given several bits of evidence that would I think show that having a fracture liaison service actively engaged in patients who are discharged from the ER to have improvement of their care and getting the appropriate BMD testing or treatment.

But what I could not find was evidence that the act of giving a discharge instruction sheet with recommendation to see their primary care physician or a specialist was – I didn't see the evidence that that was showing any change in outcomes. So, all of those pieces of evidence that were provided were actual active engagement of the patients and/or the provider, but not by giving an instruction, discharge instruction sheet.

So, as I ran the algorithm I came out with evidence insufficient because of that. Certainly I thought there was a performance gap, as we've mentioned in the last one, with only 20 percent of patients with fragility fractures ending up having testing or treatment, and no treatment can improve the risk – decrease the risk for future fractures.

So, in terms of the reliability testing, it was at patient level data and the Kappa score was 0.76, and it was state – they stated moderate to high reliability, again, with the same sample size of the previous study.

On the validity testing, it was done with electronic survey and focus groups. I couldn't find a good description of the number of participants in the survey or the focus groups, the response rate and the like. And it was a lack of – I couldn't find a statement about the safe validity assessment. So I thought the validity was (insufficient) as well. The registered validity, I didn't have any concerns about.

Feasibility, I think that it'll be fairly easy to find that a patient has been given a discharge instruction sheet, but the details of what's on that discharge instruction sheet would be a little more difficult to collect in a electronic format. I think that that would require a safe or a manual chart review to find

out that they were given instructions to see their primary care physicians or a specialist because of possible osteoporosis.

But that, again, is not – it's just the difference between electronic versus manual abstraction of the data.

Karen Johnson: Great. That was a great summary.

(Bill Curry): So, I think, yes, the use and usability, I think the issues were – I think it's easy enough for an institution to be able to add that to a set of instructions that they're going to give to patients or caregivers at the time of discharge. The only unintended consequence I saw with this was would there be a potential break in the patient's PCP relationship if there was an automated means, if they went to a fracture liaison service. And before they got back to their PCP or something, would there be the appropriate communication between the fracture liaison service? I think it's a minor concern. I thought that use and usability was reasonable.

Female: If I can just say one of the discussions around fracture liaison services has been that need for communication with the PCP for ongoing care. So, I think that's kind of one of the two missions of the fracture liaison service is the engagement with PCP.

(Bill Curry): Yes, I think that the evidence is there to support that they offer a lot to this group of patients.

Female: And I think, if I might just add, the reason fracture liaison services were developed though is because (handing a chute) to people in some smaller studies show that it didn't really work.

(Bill Curry): Yes.

Karen Johnson: OK.

Female: (Inaudible) its' a good idea about it.

Karen Johnson: So, how do you guys feel about the process itself in terms of being able to find the information that you need in the submissions and walking through the algorithms? Any questions burning for me about the method?

Female: I didn't find the method difficult. What I found difficult was when the elements are very different and the science doesn't equal for each of the elements, that was harder for me. Yes.

Karen Johnson: So, you know, well …

Female: Yes.

Karen Johnson: Yes. I think that's a very valid question to think about. The project team is pretty much available to all of you if you have any questions about these measures or others that you'll be looking at in the next couple of weeks. So, please give us a call or an e-mail if you get stuck on something and just want to run something by s), we're happy to do that.

Male: So, I mean, when will we be able to look at the comments that are – that other members of this team have submitted for these three measures?

Karen Johnson: I believe that we have given you the comments that we have received. So, I don't think there are any other ones.

Male: I'll try again. The last time I tried to interact before this meeting I could only see my own again. I'll try again. I'll look at it.

Karen Johnson: Yes, when you go in you'll only see your comments? Well, actually, there's – we're adding them into the measure worksheet. I don't know if that is where you were looking.

Male: OK, OK.

Karen Johnson: Yes. So there's a separate box for the three workgroup comments.

You know, real quickly before we sign off here, we don't want to keep you too much overtime. If you have not received an e-mail from our meetings

department regarding travel arrangements, please send me an e-mail and I'll make sure that you get that.

(Crosstalk)

Female: Can you just remind me what the next steps are for each of us with these different measures?

Karen Johnson: Sure. So, for next step, NQF staff here, we're continuing our workgroup call. We have a few more left. We'll still ask the lead discussions to be prepared to walk us through the measures at in-person meeting. So just if you could continue reviewing the measures in this workgroup and the others as well in preparation for the in-person meeting, I think that's what we're really asking of you right now.

Female: OK. Thank you very much.

Karen Johnson: And, yes. Any questions as you're reviewing them, please feel free to give us a call.

(Anne Watt): This is (Anne) from the Joint Commission. I had a question too. Will you be looking for us to do a general introduction to the measures or will the workgroup members be doing that?

Karen Johnson: We will ask you to do a general introduction. Probably for this one, (Anne), we probably wouldn't ask you to do it three times. And we may ask you to just talk about the measure set and, you know, it might be very similar to what you have shared with the workgroups.

(Anne Watt): Very good. Thank you.

Karen Johnson: OK. So, thank you everyone for joining. We hope you have a nice weekend and this will end today's call.

Male: All right. Thank you.

Female: Thank you.

Female: Thank you all.

Female: Thanks, Anne. Thanks, (Cathy). Bye.

Female: Thank you.

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

END