NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
April 24, 2014
1:00 p.m. ET

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

(Katie Streeter): Hi, good afternoon, everyone. This is (Katie Streeter) here at NQS, and I'm with (Angela Franklin) and (Anne Phillips). We'd like to welcome you to today's musculoskeletal workgroup group two call. I'd like to remind everyone that this is a public call, and I'd also like to welcome our developer colleagues who are joining us today.

Now we'll proceed and have a roll call, and we'll be calling out your name.

Female: (Kim Templeton)? (Kelly Clayton)? (Linda Davis)?

(Linda Davis): Here.

Female: (Marcy Harris Hayes)? (Mark Jarrett)? (Huja Kana)? (Jason Masusik)? (Arthur Shuna)? (John Ventura)?

(John Ventura): Here.

Female: Great. I think – we have several people that seem to be logged in on the webinar. We might just need to wait a couple minutes as they are dialing in to the conference line.

(Wendy Marinkovitz): This is (Wendy Marinkovitz). I'm also on the line. It's not my group of measures, but I wanted to listen in.

Female: Oh, no, thanks for calling in, (Wendy).

Female: Is there anybody else for the committee who – besides (Wendy) who's not in the workgroup who's on the call?

Female: I am in the workgroup. It's (Kelly Clayton). I spent about four minutes waiting for an operator to connect, so...

(Kelly Clayton): Uh-oh. So we may need to give it a couple more minutes.

(Marcy Harris Hayes): (Marcy Harris Hayes) has also joined, who's on the committee.

Female: Hey, (Marcy). Thanks for calling in.

(Kim Templeton): And (Kim Templeton).

Female: Great.

(Genuth Yazdani): This is (Genuth Yazdani) from the American College of Rheumatology, and I apologize ahead of time. I'm sort of going to be in and out today.

Female: OK.

Female: OK. You were a little faint. Could you repeat your name again? I'm sorry.

(Genuth Yazdani): Sure. It's (Genuth Yazdani).

Female: Great, thanks, (Genuth).

Female: Do we have (Mark Jarrett)? (Huja Kana)? And (Jason Masusik)?

Male: Hello?

(Huja Kana): Hello. This is (Huja Kana) here.

Female: Hi. Was that (Mark)?

(Mark Jarrett): (Mark Jarrett), yeah, hi, they needed my Social Security number, I think, to get on the line. Did you guys already do roll call or attendance? Because I just joined on, too.

Female: Hi. Is this (Arthur Shuna)?

Female: Who was that who just joined?

(Jason Masusik): (Jason Masusik)?

Female: Oh, great. Thank you.

Female: Hi, (Jason).

(Angela Franklin): I just put the phone number up in the chat. OK. I think we (inaudible) great. So welcome, everyone, to our second workgroup meeting. This is (Angela Franklin). And, again, we're joined by (Katie Streeter), our project manager, and (Anne Phillips), our project analyst. And thanks already for your hard work on the measures so far.

I wanted to let everybody know that we won't be – I wanted to go over the purpose of today's call. First of all, I just wanted to let everyone know, we won't be voting today. We're saving that for the in-person meeting. But today's purpose is to have an in-depth discussion about the measures as a small group ahead of the full steering committee meeting in May and really bring to light any specific issues that we may have with the measures. And we're lucky to have our developer on the – developers on the call to answer any questions about the measures.

And just real quickly, I know that we have the developer from (MCQA) on the call. I just wanted to check and see if we had the developer from American College of Rheumatology on the call?

Melissa Francisco: Yes, this – hello?

(Angela Franklin): Hello.

Melissa Francisco: Hi, this is Melissa Francisco from the American College of Rheumatology. We indicated that we would have me, and (Amy Miller) should be joining us, as well, as representatives from the development team. However, we may be limited in our ability to answer certain questions.

(Angela Franklin): OK. OK, great. That's good to know. And just to tee off of that, I wanted everyone on the steering committee to know this is a good opportunity to ask those questions about the measures, but if we have detailed questions or questions that the developer can't answer, we will certainly follow up via e-mail with both the committee and the developer. Also, if there are any questions at the end of our measures discussion today regarding the full steering committee meeting, we're happy to address those, as well.

We're also piloting several new aids for the steering committee during this project, including the committee guidebook and the measure evaluation algorithm. And if you could give us any feedback you have on those as we proceed, that would be most appreciated.

So just before we get started and dive into our first measure, I'd like to set the stage a bit for the measures we're going to be discussing today. We have five in total. That's one maintenance measure and four new measures that happen to be e-measures, according to the NQF definition. Because we have five measures, we'll give approximately 15 minutes per measure for discussion.

And I also wanted to mention, we are piloting in this project for e-measures only a trial implementation pathway. So just to quickly run through that, for these e-measures, staff is going to be conducted an internal review of the e-specifications. And we'll provide a report to the steering committee prior to our in-person meeting in May, so the committee does not have to review this particular aspect of these measures. And as I mentioned, NQF is piloting a new trial implementation pathway for many of the e-measures in this project.

And this means that while the committee should dig into the evidence criterion and the feasibility and use criterion, for these particular e-measures, they could be eligible for endorsement by – I'm sorry, they could be – have areas in terms of the scientific acceptability that aren't quite complete or are

not there, and these still, however, could be eligible for endorsement for trial use per our pilot. So please keep that in mind as you're reviewing the measures.

And in addition, we do have some harmonization issues today, which we won't get into at this time, because staff is communicating back and forth with the developers and will be, again, preparing a report for the committee prior to your in-person meeting.

So just a process for today will be, for each (inaudible) discussion to introduce the measure overall, give a quick description of the measure. Then we'd like to have you quickly summarize the initial comments from the group about the criterion that's in focus, for instance, going through importance. Then ask secondary discussants for any additional thoughts without repeating information. And then open the door for discussion of the criterion by the full group.

And we ask that questions of the developer be asked after the floor has been opened to the full group so we can ensure we have crisp and crystallized questions for the developer to answer. Are there any questions from the committee members about any of these items so far? No questions? OK.

With that, I guess we'll start with our first measure, and that measure 0054, disease-modifying anti-rheumatic drug therapy for rheumatoid arthritis. And the lead discussant I have is (Kuna) – and (Dr. Kuna) and (Dr. Shuna).

(Huja Kana): It's (Huja Kana).

(Angela Franklin): Ah, thank you.

(Huja Kana): That's OK. We sounded like twins there.

(Angela Franklin): I know, that's – sorry, that's my...

(Huja Kana): That's OK, no worries.

(Angela Franklin): Thank you.

(Huja Kana): So I guess – and this is (Huja) here – I can give a brief overview of the measure, and then maybe (Dr. Shuna) can take over. Is that all right?

Female: That would be great.

(Huja Kana): OK. OK, so this is, again, measure 0054. And just so you guys know, I'm a rheumatologist, so I have, you know, a vested interest in this measure. The measure title is "Disease-Modifying Agents for Rheumatoid Arthritis."

And to briefly describe the measure, what it's asking us to do is look at the percentage of patients who are above 18 years of age at the end of measurement period and are diagnosed with RA and have at least one ambulatory prescription of a disease-modifying anti-rheumatic drug.

So the numerator that we are looking for is patients who have RA and have this one prescription, and the denominator, of course, is all patients above the age of 18 and have had an outpatient visit for the diagnosis of RA or have had inpatient discharged – or, actually, and have had an inpatient discharged with a diagnosis of RA.

There are some exclusions to the denominator. We are excluding folks who have HIV and women who are pregnant, because we cannot use DMARDs, which are teratogens, you know, for these patients.

The measure is a process measure, and the data sources are administrative claims data, electronic data capture, obviously, from pharmacy. And the level of analysis was done at the help line levels for Medicare and Medicaid and – you know, on an integrated delivery system.

So from what I understood was this was originally endorsed in 2009, and this is the – we are asking for re-endorsement over here. And in reviewing the worksheet, following the comments that came from the group – and – you know, I cannot identify who made each one of these comments, other than myself, but as far as quality construct and rationale goes, it appears that all four of us agreed that, you know, the quality was high, there was, you know, a

detailed summary provided, and – you know, the reliability and validity were obviously met.

The cause is just, because it is a high-profile disease. There is a lot of mortality, morbidity concerns with it. And the HEDIS data is definitely showing a performance gap, so there is a large difference in the number of DMARD prescriptions among different plans, and that's why it underscores the opportunity for improvement.

The evidence that was provided to support the cause was also extensive and – you know, for folks who haven't looked at the citations, there are more than 20, if I'm not mistaken, that directly feed into it. And all of them are randomized controlled trials, if not all.

It's a relatively feasible measure to obtain data on, because these are all electronic data captures, so the data elements are routinely generated as part of routine care, based on medication orders, ICD-9 codes. I mean, you have to have those two things to write a prescription anyways. So it's fairly easy to do.

From a usability standpoint, it's already in use. And there are – you know, the benefits outweigh the negative consequences. There are public reporting processes in place, the Medicare Stars program, the physician quality reporting system. All of these are already doing that.

As far as competing measures (inaudible) I don't know the clear answer to that. I don't believe there is a competing measure, but I may be wrong.

Male: Well, actually, one of the other measures – which is an e-measure – is going to be – because that puts a timeframe within a year.

(Huja Kana): OK.

Male: (inaudible)

(Huja Kana): So, great.

(Angela Franklin): This is (Angela). We did identify 2525 as a competing...

(Huja Kana): OK, OK. I'm sorry. I did not review that, so I wasn't aware. So that's kind of the brief summary of it. Was there any other detail that you would want me to provide?

(Angela Franklin): No, this is great. And if we have the secondary discussant, we can have his thoughts. (Dr. Shuna)? No.

At this time, we could throw it open to the floor for discussion of the importance criteria.

(Mark Jarrett): This is (Mark), also a rheumatologist. I mean, I think the literature is pretty clear about the need to put people on DMARDs. You know, the exclusions seem appropriate with the HIV and, obviously, the pregnancy during the measurement year. The only thing that could theoretically be in there is there are other patients with other immuno-compromised issues, patients who perhaps, let's say, have had a history of malignancy, and some people will be afraid to use certain DMARDs in malignancies. So that might be the only thing that I would raise as a possible issue.

(Jason Masusik): This is (Jason Masusik). And then taking a look at this measure and the competing measures, one of the other groups that the competing measure had raised as a possible exclusion, as well, was the patients who maybe were in clinical remissions or didn't really – or patient preference, you know, if there's exclusions made for those types of situations, as well. And I think that the quality – the evidence that they presented, though, was excellent.

(Linda Davis): So this is (Linda Davis), and I am representing a business coalition and not a clinician. So please be patient with my questions, but as I understand it, there are multiple DMARDs, and some are biologics and some are not. Is that accurate?

Male: That is accurate.

Female: That is correct.

(Linda Davis): So this includes all of the DMARDs, regardless if they're biologics or not. And it's my understanding that biologics in many cases may have their data in – this might not be relevant to the topic that we're talking at this very moment, but their data is in both medical claims and in pharmacy claims. So I'm curious about just the complications that might present in that case.

And then I'm also wondering about whether – what is the goal of this? Is more better? Is a higher number, a higher percentage of patients a better number? Or what is the conclusion that we draw when we see variation in these numbers, perhaps?

(Huja Kana): So let me answer part of that question for you. So the agents that are included – yes, you're right – are all oral (inaudible) biologic and non-biologic agents. As far as, you know, whether – you know, more is better or not, that's not actually true. What is better is early treatment of RA. That is the whole rationale for this measure, that the earlier we start the disease-modifying agent, we are trying to halt that inflammatory process ahead in time and stopping the progression of disease. So that is the true rationale behind this, you know, measure.

(Linda Davis): OK.

(Huja Kana): Did that answer your question?

(Linda Davis): That answers...

(Huja Kana): Part B.

(Linda Davis): Yeah. Yeah, that's fine. Thank you.

(Kim Templeton): And this is (Kim). If I could ask a question – I'm sorry if this is too silly, but what we're looking at are the number of – in the numerator, the number of patients who received a prescription or those who filled a prescription?

(Huja Kana): Who've received an ambulatory prescription, yes.

(Kim Templeton): OK, wonderful. Thank you.

Female: So that may not – that may mean they didn't get it filled potentially.

(Huja Kana): That is true. That could be a possibility, because we also have to take into account patients who, for instance, you know, get pharmacy benefits from another plan. And, you know, this is not – you know, this is not their primary plan, where the performance measure is coming from, or, you know, they're paying out of pocket, for instance. So we may be missing some data over there.

(Jenna Williams Bader): Hi, everyone. This is (Jenna Williams Bader), assistant director of performance measurement here at (NPQA). I actually wanted to jump in, if that's OK, and address the two points that were made, just to make sure the workgroup is clear. This measure actually looks at whether patients were dispensed a DMARD, not prescribed a DMARD. So we are actually looking for – right, for the actual dispensed drug.

And then, also, the – when we use this measure in our HEDIS – in HEDIS, in our accreditation program, we actually require that the plans for reporting it, that the patients they include for the measure have both medical and pharmacy benefits with that plan. So they will not be receiving their pharmacy benefits from another plan.

(Huja Kana): Perfect. Thank you for clarifying that, because that's where I was having trouble.

Female: I guess that would be concern, then, is that you can't always control what patients are going to do, and so they may receive a prescription, but if they don't fill it, that's not necessarily a reflection on the quality of care that they received.

(Jason Masusik): This is (Jason) again. So should those count as exclusions in the denominator that are not listed here within the application, if they don't receive medical and pharmacy benefits from the same – the same provider?

Female: That sounds like a question for the developer as to...

(Jenna Williams Bader): Yeah, that's actually a good question. When we submit measures to NQF, we're actually instructed to be somewhat program agnostic, which does make this kind of situation a little tricky. The – we consider the requirement of both medical and pharmacy benefits a program requirement for plans that are reporting this measure as part of our accreditation. So it's just a fine line about deciding whether that's a part of the actual measure's specification or requirements of the program.

(Mark Jarrett): This is (Mark) again. I apologize. It may be answered and I just missed it. What about patients who have inactive rheumatoid arthritis, maybe have some secondary osteoarthritis, are on some non-steroidals, you're fine, but you're going to carry the diagnosis of rheumatoid arthritis, especially in the world of electronic medical records, where everything gets carried forward, and also as people are trying to demonstrate the comorbidities in their patients in terms of risk adjustment and everything else with – you know, in the world we live in.

Obviously, a patient with inactive disease, which, you know, in practice you do see, do not – you know, go into a (natural emission) – especially early in the disease may happen. Prescribing it would not be appropriate. They'll be in the denominator, but they're going to throw off the numerator.

Female: My understanding is that inactive rheumatoid arthritis would be difficult to capture via claims. I'm certainly interested if there is a way to capture it in claims. It's not something that we would be able to incorporate right now into the measure, but it would be something we could perhaps pursue in the future, if there is a way to capture it.

(Mark Jarrett): Yeah, I don't know if it maps out in ICD-10.

Male: Well...

Female: As it was explained earlier, this goal is – the goal of this measure is to – well, to – the appropriateness is – earlier on is to increase utilization early on in the onset. And is there a way through claims data to know if it's – what stage of the disease the patient is actually in, if it's early, if they've had it for a long time?

(Mark Jarrett): Well, that gets back to 2525, which is a prescription for a DMARD within the first year.

Female: Right, that's – yep. Yep.

(Mark Jarrett): So the question is, do we – and I'm just asking – do we need both? Or is the other one the more important measure, because this is going to have that denominator containing the whole population, which is including inactive, active, again, people with other comorbidities, why maybe they're not going to be prescribed. And I just worry that I don't know what this value is really going to mean.

(Angela Franklin): So this is (Angela). That's an excellent question, and it's part of what we'll be doing and discussing with developers as we go back and forth with them regarding the harmonization issues, so we will definitely include that as part of our question.

(Mark Jarrett): Thank you.

(Angela Franklin): I just – with an eye on the clock, this is our importance criteria for this, and I wanted to say, if there's additional questions or concerns about importance, if we could raise those now, if not, move on to the next criterion, which is scientific (inaudible)

(Jason Masusik): I guess I just had one last thing. This is (Jason) again. In taking a look at the data that looked like it was presented, if I'm remembering this correctly – I don't have it up on my screen right now, but it looks like over 90 percent of commercial plan patients were actually meeting this measure, whereas Medicare and Medicaid were slightly lower rates. What is the threshold that NQF looks for, for meeting, in terms of, you know, is it still somewhere that you want to spend a lot of resources? And if 90 percent of commercial plan payers are already meeting this measure, obviously, Medicare and Medicaid has some room for growth. But do you have anything internally? Is it the 90 percent mark that you really look at? Or is it something different?

(Angela Franklin): It's something that the – I mean, sorry, the committee kind of looks at in terms of – I mean, in terms of perhaps being popped out as a measure. Is that what the question is?

(Jason Masusik): Yeah, yeah.

(Angela Franklin): It's really something that's the judgment of the steering committee. We don't have a hard and fast threshold. And as you said, there's some room for some of the other carriers to improve. So that's definitely a topic for discussion by the committee. Is there any other questions about that or discussion about that?

Well, hearing none, if we could move on with our primary discussant to scientific acceptability.

(Huja Kana): Sure. So going over the specification, the group thought that – I think that from a reliability testing, the overall reliability estimates were like 0.87 for the commercial plans. Medicaid was 0.89, and Medicare was 0.93. And this was out of 355 commercial plans, 98 Medicaid and 383 Medicare plans. So it was considered pretty reliable, not only for overall, but for individual, as well, because the reliability estimates were above 0.7.

There was a comment about the performance measure scores as far as the validity being limited because of the accuracy of claims data. I don't know if anybody wanted to discuss that further.

Male: I think that just goes back to what we've already talked about, that sometimes claims data doesn't give you a whole picture, you know, especially since we weren't aware when we were reviewing this measure that it was going to be limited to patients that were only be dispensed meds through their medication plan that was associated with the carrier, you know? So people were paying pocket for – paying out-of-pockets for meds. How do you capture that in claims data? And, you know, just notoriously claims data isn't always the best way of extracting data, so...

(Huja Kana): Right. And I hear that point, because, you know, my concern also was the fact that the agreement between the administrative data and the medical

records data was also low when done in the field testing. It was like close to like 57 percent or so. So that definitely stays an issue. Do we have any other thoughts or should I move forward?

(Angela Franklin): I just wanted to do a quick check to see if the developer had a comment about the field testing.

(Jenna Williams Bader): Yeah, hi, this is (Jenna). I guess I would say, as far as the validity goes, we do actually – there is quite a lot of variation in the plans. We only – we typically field test with only three plans, because it is quite an endeavor. So actually when you look at the results from that, the validity varied – did vary quite a bit between plans, and there were some where it was – the rate of agreement was actually quite high, and others where – there was one in particular that was below 60 percent.

So I think it's somewhat difficult to judge based on just those three plans, how valid this information is. But I think that's really all I – I can answer other questions, but I think that's the best point to make about the field testing.

Also, I guess to go back to the questions about patient paying out-of-pocket, there are some instances and measures where we know that we might be missing something and not picking up every single patient, but we assume that the distribution of those patients, types of patients would be equal across plans. So this could be a situation where if patients are paying out of pocket, we do assume that there's going to be fairly even distribution of that among the different plans or cost of plans.

Male: But is this measure designed only to measure the quality that the plans are generating or individual physicians?

(Jenna Williams Bader): This is a plan-level measure, which is one bit – a significant way that it differs from the next measure that you'll be discussing, 2525, and it's implemented at a plan level in quite a few different programs.

Female: I don't know about the assumption that all patients – all plans will have the equal number of patients without with out-of-pocket coverage, since we know that high-deductible health plans vary by geography. They're all increasing

as, you know, the time passes, but there's quite a bit of variation across the country because certain states and their penetration of high-deductible health plans.

(Angela Franklin): Thanks. Are there additional questions about the testing, scientific acceptability? Comments? Well, hearing none, we can move on with the lead discussant to the next criterion.

(Huja Kana): So the next criterion is feasibility, which looks at the – you know, the extent to which the specifications included – including measure logic, required data that are readily available, or could be captured without undue burden and can implemented for performance measurement.

And it seems that we've already pointed this out, that this is based, you know, on – it's a plan measure – the data is being captured electronically. It's easy to get the diagnosis codes through the ICD-9 codes. You know, if the prescription is given and dispensed, it's all electronically available. So it appears that it is feasible to obtain this data, and it can be corroborated through medical records, as well.

Let me open this up and see if anybody has any questions about that.

(Linda Davis): This is (Linda Davis) again. And I believe, though, that unless things have changed in the last couple of years, prescription data does not necessarily have diagnosis codes on it. That may have changed, but I think this does require that you tie the medical claims data patient to a prescription file that is an entirely different data set. And so it does require some mapping between the patient's medical and pharmacy claims data, correct?

(Jenna Williams Bader): That is a good question. I cannot answer it right now, but we'll certainly look into it and provide an answer to that question.

(Angela Franklin): Are there additional questions or comments about feasibility? OK. Hearing none, I guess we can move on to the next criterion.

(Huja Kana): So the next one is usability and use. Looking at the extent to which the potential audiences, (so consumer purchases) providers and policymakers are

using or could use this – the performance results for both accountability and performance improvement to achieve the goal of high-quality efficient health care for individuals or populations.

And from the comments that were reported, it seems that this measure appears to be used widely in – across multiple plans and rating systems for health care quality and – you know, it seems that it has demonstrated improvement in performance for a number of health care plans from 2010 to 2012, but there are areas that need improvement.

So, you know, and I think I'm the one who put in all the places where it's being done, so Medicare Stars is looking at using it. There is public reporting, you know, based on the health care ranking. PQRS is looking using it, and so are the HEDIS measures.

So I think the measure gets to quality improvement and benchmarking, definitely. So I don't know if anybody else had any other thoughts where we could see improvement. I guess not.

Male: No. No.

(Huja Kana): So moving on to criterion five, I guess I stand corrected. 2525, the next measure, is a competing measure. And (inaudible) we will moving onto that next.

(Angela Franklin): Yes, we will. Thank you so much.

(Huja Kana): Sure.

(Angela Franklin): Our next measure is measure 2525, and we do have, I believe, (Mark Jarrett) – OK, thank you, (Mark).

(Mark Jarrett): OK. I apologize. I (inaudible) conversation, because I thankfully had a week off, but it was in the rainforest of Costa Rica without very much Internet access, which was very pleasant, let me tell you, for a week.

But, anyway, going through in terms of, you know, the measure, this measure differs from the other measure, that it's basically a percentage of patients who

are over 18, diagnosis of RA, who are (inaudible) disease-modifying drug within 12 months of their diagnosis. And it's an issue with exclusions of HIV, again, patients who are pregnant. This one does have the exclusion of inactive rheumatoid arthritis.

If I am not mistaken, again, reading through while I was away, that this is really an e-measure that measures, really, a similar one that was done not as an e-measure in the past. Am I correct?

(Angela Franklin): And I would throw that to Melissa Franco?

Melissa Francisco: Yes, that is correct.

(Mark Jarrett): OK, so it's basically getting it out as an e-measure, and I'm not going to go into how you're going to pull that out. We'll talk in the feasibility part. So, basically, it gets at the aspect of early treatment rather than just somebody being on the drug, and to the issue of inactivity of disease, one can assume that in the first, you know, year of treatment, first year of disease, most likely the patients will have active disease at that time.

Again, I bring up the – in the denominator issue of malignancies, because except for one or two DMARDs, people are going to be very hesitant, especially with the biologicals, to give it to somebody who's had a recent or active malignancy, because of the interaction either with the immune system or the interaction with the drugs themselves, if they're under treatment at the present time.

So that's – that's the measure in a nutshell. And I'll take from anybody else in the group who has anything else that they think about it.

(Jason Masusik): One of the – this is (Jason Masusik) – I had a question about the – about what it means to be diagnosed. Is that, you know, by the patient's primary care physician, but they don't get in to see a rheumatologist for six months, and then, you know, is it two visits from that point or – or 12 months from that point? Or is it 12 months from the time that the initial doc put down the first ICD-9 codes for rheumatoid arthritis? Or is it 12 months from the time that

you get the first positive laboratory results back? Or I don't think I was fully clear on that piece of it.

(Mark Jarrett): To me, reading through, it's not clear, either. And we have to be careful, because there are – especially in areas where there aren't a lot of rheumatologists, there are primary care physicians who will start disease-modifying drugs, especially some of the non-biologicals, and therefore I think it has to be left often to the – to the time the first diagnosis is made, when the diagnosis is first made at that point, because otherwise they'll become too complex to figure out specialist versus non-specialist.

I don't think, you know – clearly, one could argue that, you know, if it takes six months to see a rheumatologist, perhaps – and that primary is uncomfortable using these drugs, perhaps they have to get them to another rheumatologist, because, you know, it's much like anything else. You're not going to send somebody, give them a slip for mammography, and because you feel something and they can't get an appointment for three months, and just say, well, we have to live with it.

So maybe leaving it this way may not be a bad thing, because it will – you know, it will prompt more rapid closure on getting them an appointment, if that's what's necessary. That's just my opinion.

(Angela Franklin): And this is (Angela). This might be a point that Melissa might be able to address?

Melissa Francisco: This is definitely a point that I can take back to the group.

(Mark Jarrett): Yeah.

Melissa Francisco: As I think will probably be the case for most questions that you're going to ask today.

(Angela Franklin): Thanks. Other comments on importance?

(Mark Jarrett): Do you want me to go on with the quality construct at this point? Or...

(Angela Franklin): Sure, sure, please.

(Mark Jarrett): ... a lot of comments on that, which I think were important in – and that gets to availability of rheumatology, you know, and one of the comments that's brought up about socioeconomic status and insurance status. Clearly, since, you know, this is coming from the ACR and not – you know, American College of Rheumatology – and not NCQA, the question comes about availability of prescriptions and cost of prescriptions, et cetera, which gets to the bigger issue that I know NQF has now, you know, sent out the comment period on about risk adjustment, about how we do it, the socioeconomic factors, and my concern – and I would bring it back to ask ACR to look at it – is, is there any way they can think that they can have somehow some kind of modification, because results may be low, not because either the patient or the doctor are not doing their thing, but the patient, quite frankly, can't afford the drugs.

(Angela Franklin): Any comments regarding quality construct?

(Mark Jarrett): And if I can make one more, that – somebody else commented, they said it was level C evidence, you know, because it's based on guidelines, not randomized controlled trials. But at least from my viewpoint – and I'd be glad to hear what the other rheumatologists on the call have to say – you know, I think it's still part of the mantra and it's considered the standard of care today.

So I would find it hard to do – you know, to see somebody doing a randomized control study at this point, looking at this – at whether you start people within the first year or not, but I'd be – I welcome comments from the others on the phone.

(Huja Kana): This is (Huja) here. I would agree with that. It's going to be very hard to pull off at an RCT looking at this. I think the clinical guidelines that we have currently are based on randomized control trials, so we have plenty of data. I don't think we have to beat around the bush anymore. We just have to go ahead and start treating people more aggressively.

(Mark Jarrett): I agree.

(Jason Masusik): So that would bring the next point up – this is (Jason) again – in that looking at the – actually, if we back up for a second to the importance criteria – I'm sorry to go backwards on that – but it looks like the ACR clinical registry found that 99 percent of patients in 2012 in the registry were actually already achieving this measure.

So has – if everybody's already doing it in all of rheumatology, because it's the standard of care or the de facto standard of care or the – what everybody has decided should be done, does this need to be a performance measure? Should – has this reached a point at which there's any potential for improving that further?

(Mark Jarrett): And I may be mistaken about this, but not everybody is in the – I mean, I don't think the registry demonstrates – you know, reveals – is representative of the denominator of people with RA within the first year of their diagnosis in the United States. So I think you're looking at the people who participate in the registry are – you know, are the people who've gotten with the program, so to speak, and the real question is, what about the rest of the country?

So I think that measuring this may be valuable. But, again, I defer to the others on the phone what their opinions are.

Melissa Francisco: Hi, this is Melissa, and I can speak to the point about RCR data and participation. (Mark), you are correct. It is a limited group of rheumatologists who do report through the registry. So you're looking at a very narrow subset of rheumatologists when looking at these data. And they are typically people who are more with the program, who are looking for quality improvement and who are already on board with that.

Female: So, Melissa, these are folks who are coming from academic institutions, if I'm not mistaken, correct?

Melissa Francisco: No, these are actually clinicians.

Female: They are clinicians in the community?

Melissa Francisco: Yes.

Female: OK. So, (Jason), to answer your point, there is a gap we still needed as a performance measure.

(Jason Masusik): Yeah, and point taken. I was just looking at some of the information that was presented, and it just didn't make sense that, you know, where they're presenting data and showed a very, very high uptake rate, and (inaudible) physical (inaudible) so...

Female: Yeah.

(Angela Franklin): Are there additional questions under importance? That would the importance, gap, priority or quality construct. OK. I think we can move to the next.

(Mark Jarrett): This is on the scientific – this is on the scientific acceptability and what we – some of the comments were – you know, everybody felt things were high. However, there was a concern – and I guess, I don't know how much we need to comment on this in terms of – because I know you said you're going to have a separate e-measure group looking at this.

But there is one area where the agreement between the e-measure and chart review, you know (inaudible) was about 0.67, which is not bad, but it's not great. And the question is, will the e-measure itself be different than what we're seeing on paper? And, therefore, are we at a time with the EHR, you know, for the next year that we feel an e-measure will really reflect what's truly going on. That's, I think, some of the comments that I saw that had some questions.

The other question was the date of disease onset. You know, someone may come in to see the primary – even their primary six or nine months after self-treating themselves with anti-inflammatories, the primary does the work-up, it's almost the end of the year already, and if you're doing that, so, really, what's the – what are we using as the benchmark for the onset of disease? So those are the two questions that were raised, really, the major questions by the workgroup.

(Angela Franklin): So this is (Angela). I just had a quick question for Melissa about the specifications about determining the date of disease onset, if there was any clarifications you might be able to let us know about.

Melissa Francisco: That's a question I can take back to the group, as well.

(Mark Jarrett): It came up in feasibility, as well, so – you know, we'll kill two birds with one stone.

(Angela Franklin): Any other questions on scientific acceptability? So hearing none, I guess we can move on to feasibility criteria, and we've already, I think, earmarked one question.

(Mark Jarrett): One question. Another one was, you know, patients, if they're getting biologicals, they're going to be tested for TB, and that might be a contraindication, because they might be positive and need to be treated. And the question is, are we going to be able to capture that information? Or, really, should we be adding as an exclusion a positive tuberculin test requiring treatment? Which may be difficult to figure out which of those patients, but just asking that, because, again, you know, you may have to wait 6 or 12 weeks to treat the patient before you start him on the drug, and that might be an issue.

(Angela Franklin): Were there additional questions about – or comments about feasibility?

(Mark Jarrett): I mean, the feeling was – my feeling also was that if we get the e-measure right, it's better than, you know, using claims data, which always stinks.

(Jason Masusik): So just a question on this from (Jason) again. Now, just so that I'm clear, is this that they're prescribed the – the disease-modifying drug or that they actually filled the prescription and started taking it within the interval period that you're looking at?

Male: OK (inaudible)

(Mark Jarrett): I'd assumed it was prescribed, because I couldn't figure out a way that they'd be able to gather the data on, you know, picking up the prescription, since these patients have all or no kinds of insurance, and it's not through NCQA.

(Huja Kana): Well, it states newly prescribed...

(Mark Jarrett): Yeah, that's what I...

(Huja Kana): ... but it doesn't say dispensed.

(Mark Jarrett): Yeah. And I think it really needs to be prescribed.

(Jason Masusik): Yeah, well, I think in order for it to be an e-measure where we're actually able to capture the data, it sounds like that's probably the only way we'd be able to do it, is just based on whether or not the prescriber actually physically prescribes the drug, whether or not the patients pick it up (inaudible)

(Huja Kana): Correct.

(Angela Franklin): OK, I thought I heard maybe another question for the developer around clarifying, or if that's even possible, just (inaudible)

(Genuth Yazdani): This is (Genuth Yazdani). I'm sorry, I have to go in and out today, but I think I can take this one. We actually have three ways of capturing the DMARD prescription, one (is it appears) on the active medication list, one is it appears as administered, as would be the case with some biologics, and the third is whether the medication was ordered, so we use the entire EHR to the best of capacity to capture this.

(Mark Jarrett): So let me throw this out there, since if you wrote a prescription once, and the patient never filled it, you'll get – you know, the doc, so to speak, gets a positive plus check, but in reality, if they come back to the doctor and the doctor knows they didn't fill it and tries to convince them again, and writes a second prescription, I hate to say that shows more due diligence than – you know, it's – again, I'll use the example of mammography. Just because you hand somebody a slip doesn't mean they go for it, but if they come back to

your office and say, why didn't you go for it or you follow up on it, that's really what you need to do. You can't get 100 percent compliance.

So I'm just going to throw out there the possibility of saying, if you're using just prescribed and not actually filled or actually giving like a biological from a claim, then I would say – throwing out there the thought of two prescriptions, just throwing it out there to show that, you know, if we're really trying to find the people who really aren't doing it, I don't know if it's feasible. You know, I don't know if there's – you know, if it's better or not, but I'm just throwing it out there as a thought.

Female: The question gets to the issue of adherence, which I think is an incredibly important concept and one that sort of deserves its own measure. The challenge here is, especially with the different DMARDs – let's use the example of rituximab, which is something that somebody might just receive...

(Mark Jarrett): One.

Female: ... right, exactly, sort of – you know, a...

Female: (Single infusion), yeah.

Female: ... infusion of, and then that infusion would last for, you know, months and months. And so I think it's difficult to quantify the number of DMARD prescriptions, given the heterogeneity in how these things are dosed. And, unfortunately, we don't have great linkage between most EHRs and prescription pharmacy data, at least right now we don't, and so that's a problem.

(Mark Jarrett): Just – I was just asking.

Female: Yeah, I mean, that would be the best way to go about doing it, corroborating it with, you know, dispensing from a pharmacy, but if we don't have a way, we don't have a way.

(Angela Franklin): Additional comments about this criterion, feasibility? If not, we can move on to the next.

(Mark Jarrett): OK, so next is usability and use. Basically, the comments were that, you know, the ACR's already using the registry and, you know, it will be – I think it will, you know, be good information that people will be able to use to try and improve care, because where there are gaps, it will be important. So I think the group felt that, you know, it was positive in that regard, the comments.

(Angela Franklin): Any comments from the group on use for this measure? Nope, I guess we can move on then. Again...

(Mark Jarrett): (I think competing) we already addressed because of the previous one and figuring out how that – how the two of them fit together or maybe can be combined or something.

(Angela Franklin): Thanks, that's a good comment. Any other thoughts about that? OK, hearing no additional thoughts – and you can also e-mail staff if you have questions, and we'll get that to the developer – we can move on to the next measure, which is 2522, rheumatoid arthritis, tuberculosis screening. And I think we have (Kim Templeton) and (Linda Davis)?

(Kim Templeton): Yes, thank you. And so this measure is to look at the percentage of patients over 18 newly diagnosed with rheumatoid arthritis who have documentation of TB screening performed within 12 months of receiving their first course of therapy using a DMARD.

The incidence of tuberculosis, obviously, is quite low. However, there is data that would indicate use of DMARDs could re-activate latent tuberculosis, which would have significant – which would have significant consequences, and so this would be sort of the importance of the measure, in that – that rheumatoid arthritis, obviously, is a high-priority disease, although be a low incidence of TB. If there are ramifications to treatment, those ramifications would be significant.

In looking at the evidence, guidelines indicate that – the ACR guidelines indicate that the TB tests should be performed. Most of this, though, appears to be expert opinion, not a lot of data, although this would be something in which attempting to do a study such as in RCT would be, obviously,

extremely challenging and fraught with ethical issues. So we really don't have that kind of data, but the recommendations do exist in the guidelines.

As far as the – any sort of gap in current treatment, the most recent data that we have from the ACR registry from 2012 indicates that almost 93 percent of physicians are already doing this. So I guess that would be a question. A question that was raised by the group is, if there's already 93 percent compliance, is there really room to make significantly more improvement in that? And then as far as disparities data, not a lot of data on that. However, with this condition, with TB, there obviously are – is potential for significant health care disparities.

So with that, I'll open that up to any of the other discussants.

(Angela Franklin): And, (Linda), did you have anything to add?

(Linda Davis): Well, this may (sound ironic), but, again, being a non-clinician, I read that (there's biologic DMARDs). And so does that exclude non-biologic DMARDs? I'm assuming that it does, given the definition, but I just wanted to clarify that.

Female: Yes, that is correct.

(Linda Davis): OK. No, I don't have anything to add.

(Jason Masusik): This is (Jason). These are – actually, these next three I kind of struggled with the first section on each of these next three measures, I think, because – while there's certainly some stuff in here that sounds like very, very good clinical practice or good judgment or good thoughtful planning by clinicians, there really – and there was no evidence presented for this that – from looking at the guidebook, the committee guidebook, it really kind of indicated that there should be a distinction made between measures that we endorse based on the quality of the evidence provided, not on whether or not something was good clinically to do, and that's where I think I struggled with this one.

(John): I -- this is (John). I agree with that comment and it was -- and these measures where -- it didn't seem that there was evidence specifically on the process.

There were -- there was evidence about elements of the process, but specifically about the process itself, that following the process itself would improve the outcome of care ultimately.

Male 2: (Inaudible).

Male 1: Yes, I definitely hear with that, (John), that...

Female: Other comments about the evidence? Are there other comments about the other parts of the importance to measure and the port criteria? That would be evidence performance gap, the priority and the quality and priority.

Male 2: No, I have no other comments other than being on what was written.

Female: Great. All right. We've recorded those, and we can move onto the scientific acceptability criterion.

Female: We talked about the liability and validity of this. It's challenging because of this (inaudible) testing. It could certainly be something that is ordered by the -- by the physician or intelligence to starting treatment, but this is something that may be measured in a variety of other health care settings.

And so, I guess, as was written in the comments and in my reading of this too, I'm concerned about whether or not this measure is actually going to be a reflection of quality of care because patients may be getting T.B. testing, but if it's not at the point of care where they're receiving their (DMARDs), is this going to be able to be recorded in their -- in their records?

This could be something that's being done by their primary care physician or being done at work. There's a mention in about this, you know, being potentially -- the information being by patient self-report and then recorded; is that really sufficient and accurate?

I think this is going to be difficult data to obtain, difficult data to extract from the records, and so patients may be getting T.B. testing, but there may be no validation of this and no place to put this in the records.

So I'm not sure that in and of itself whether or not someone has a T.B. test ordered as a reflection of the quality of care that they're receiving, and so I struggle as to, you know, the people that respond on the -- that have responded to this on whether or not this really produces something that's going to improve quality of care.

That's (inaudible) turn it to any other comments.

(Jason Masusik): This is (Jason). I guess the, you know -- but I think that was the big thing that came through on the -- their survey, the ACR survey results too where the two out of the three sites or three out of the three sites indicated that they weren't sure that it -- that when they translated it to actual practice that it retained that same value in terms of deciding whether or not as a quality measure and that, I think, raised a little bit of concern for me, too.

Female: Right.

Female: Are there any comments from the developer regarding the issues raised by the committee, I think, around describing versus dispensing and testing?

(Janeuse): This is (Janeuse). I think we will definitely take this back and discuss it more fully. I think one thing that I can say today is that in terms of the consensus process of measured development, there was strong consensus that the physician who's actually prescribing a biologic (CMARD) is responsible in terms of patient safety in making sure that the T.B. testing information appears in the medical record so that even as it's done elsewhere that, you know, if you're writing the prescription that that patient safety is sort of in your hands.

So that was the rationale for the measure even given some of the issues with how data's captured currently in the EHR. We can discuss this more fully as well as (inaudible).

Female: Great. Any additional comments regarding scientific acceptability? That moves us on to feasibility.

Female: And so this is something we're, you know -- the test (inaudible) test for T.B., this should be relatively easy to order, but again, if these tests are going to be done somewhere else, the data is coming in from somewhere else, this may make it more challenging in finding how one then is going to document in an EHR.

Is there going to be separate areas where this could be documented, if it's going to be in the clinician's note? How is that data extracted in order to be able to assess for this measure. So I guess achieving the test is not the hard part. It's what you do with the information especially if it's not obtained in your facility, and how does one document and maintain the -- maintain those records so that the data can be easily extracted, and then make it -- may make it more challenging and may need some modifications in either how the data's extracted or in the EHR that the physicians are using?

Female: Other comments about feasibility? Are there comments from the developer about the data extraction or modification of EHRs needed, that might be needed?

Female: I don't think so. We can take that question back and we can get back to you on it.

Female: We can move on to (inaudible) usability.

Female: And I would welcome anybody else's comments on this because this scenario I'm not sure exactly what it is that we're looking for.

(Angela): This is (Angela). We do ask that the developer provide a plan for use and they have provided a plan, and we also want to know about any unintended consequences in the case of this new measure that committee members may have identified or that have been identified by the developer.

Female: So based on this, I mean, this would appear to -- they do have ACR (inaudible). This would seem to (inaudible) the usability criteria. So there are ways of publically reporting this.

(Angela): Yes, I would just confirm with the developer.

Female: Yes.

Female: I'm sorry. I'm not understanding the question exactly.

Female: Was the question that the measure is currently being used in some capacity already? I know that we do have a plan for use in the future.

Female: Right. That -- is it currently being reported or being used?

Female: So I'm not the registry person, but I will take the question back and I will find out exactly to what extent it's being used.

Female: And (Amy), I can -- I can -- I can jump in here. The -- there's an analogist PQRS measure that is currently in use in that program and is contained within our registry.

(Jason Masusik): This is (Jason). I had a question about the -- I see that there -- you're incorporating into the registry, but admittedly, you have said that the -- I kind of guess this goes back to developers -- you said that there has been uneven uptake of the actual use of the registry by providers across the country.

Is the target timeline feasible given some of the difficulties that you've had already getting physicians on-board with this national registry program?

Female: I think that's a -- that's an important question and certainly the ACRs are doing everything that it can to further develop the registry and increase uptake. And I -- and I will point out that uptake of the registry has increased each excessive year that it's been in use over the last four or five years.

So we have cautious optimism, but that's about all we can say at this point.

Female: Thank you. Are there additional questions? OK. That would move us to the next criterions. There are no competing measures for this one. OK. Unless there are additional questions about 2522, we can move on to 2533.

And I believe our lead (discussant) is (Christopher Viscoe). A secondary (discussants) are Dr. (Clayton) and (Matt Cusack). Do we have Dr. (Viscoe) on? If not, I would ask if (Huja) or (Marci) would be able to lead discussion.

(Marci): This is (Marci). I can lead.

Female: Thank you.

(Marci): So this -- a description of this is the percentage of people 18 years of age or older with a diagnosis of rheumatoid arthritis that have had greater than 50 percent of their total outpatient visits within the measurement year where they've had an assessment of disease activity using a standardized measure.

And so just numerator statement, again, is just this percentage of patients with the diagnosis that have had just disease activity using a standardized measure obtained greater than or equal to 50 percent of their visits.

There's no exclusions to the denominator and should I go ahead and move on to the importance?

Female: Yes, thank you.

(Marci): So with the importance -- the same question came up that both (Jason) and (John) had mentioned earlier, and probably is an important one because it really affects whether we move on from this first step to the second step, and it's in judging the evidence based on are we judging the process, which it doesn't appear like there's significant evidence to support the process; although, there seems to be sufficient clinical evidence to suggest that this would be a good clinical practice. So I think that's an important factor to consider.

Female: Great. Thanks. Are there comments from the rest of the committee?

(Jason Masusik): This is (Jason). Yes. No, I whole-heartedly agree with what you already said, you know? The evidence that they cite in here is excellent evidence, but it's all about tight control or clinical remission rather than anything showing that

actually performing the assessment greater than 50 percent of the time actually results into tighter control or improved outcomes or...

(Mark): This is (Mark). I agree 100 percent with the comments.

Female: So I guess the question is will there be further guidance for the committee in how to address this? Go on. Sorry. Was there...

(Jason Masusik): The way that I (inaudible) this was that we should, you know -- it's still good for us to finish discussing all these things even if we don't necessarily feel like the evidence is there because I think that, you know, discussing it helps us to flush out some of the other quality things, but...

Female: That's exactly correct.

Female: So then we'll just move on and hopefully this will kind of help -- we'll get further guidance on this in the future for the on-site meeting?

Male 1: Right.

Female: Yes. So in terms of guidance, we do ask that the committee look at, you know, the measure focus and be sure that that matches up with the evidence presented. So that's the guidance. And this discussion here will help enrich the discussion at the end-person committee.

So it's important that we continue to discuss the measure and any other issues we can surface about evidence.

Female: OK. That's helpful to know that we should just move on. All right. So the next area is in the performance gap and I -- there was a little bit of discrepancy in the answers from the committee and it may be related to just percentage on a percentage on a percentage.

So my interpretation from the -- from the numbers is that it appeared that there was a performance gap with some of the facilities that were measured was as low as 35 and then up to 61, which I would suspect that we would want even higher than that, but I'd be happy to hear the other members of the committee in how they interpreted that.

Female: I mean, I agree with you. That's what my interpretation was as well, that there definitely is a gap.

(Jason Masusik): I interpreted it that maybe the bar was set a little bit too low in some regards of the target was 50 percent and several of the sites or two-thirds of the sites were already well over that significantly enough so that it hold up the overall percentage and pulled that third group up -- you know, the average up over 50 percent too that maybe it wasn't a high enough target.

Female: So the -- but the 50 percent target is the -- refers to the number -- I'm asking this for clarification. So the 50 percent target in the description of the measure refers to the 50 percent of the total number of outpatient visits; is that correct? Not necessarily 50 percent. It's that criteria has met 50 percent of the time?

(Jason Masusik): Oh, I see. So that number here is more of the...

Female: Yes.

(Jason Masusik): ... percentage of practices that are at 50 percent of the outpatient visits? I understand now. Thank you.

Female: OK. Anything else on performance depth? OK. So then to look at priority -- high priority, it seems that there was support that this was a high priority measure. There was some questions, one I'm just going to read verbatim.

It's not clear to me how 50 percent of patient visits with disease activity measurement was determined. Also, is there no data provided to show these assessments are superior to those without?

Female: And if there's additional comments to respond to that from the committee, that'd be great. Otherwise, if maybe the developer would speak to that a bit?

Female: Yes, it's -- (Janeuse), are you still on?

(Jeneuse): Yes, I'm still on. I can -- I can take a crack at it. So I think one thing that's important for the committee to consider is that unlike some disease like, say,

diabetes where there is an immediate outcome that's a laboratory value like Hemoglobin A1C, in rheumatoid arthritis, really the key disease outcome is this concept of disease activity and (inaudible) experience global standards for measuring these concepts or these composite indices, which we can talk about in more detail in the in-person meeting.

And so really the rationale here is that clinicians capture the key outcome that a majority of visits in patients with rheumatoid arthritis, and the number of 50 percent was rather than being based in scientific evidence, which I think is lacking to support that specific number, was actually based on, I think, very good clinical rationale that there may be completely reasonable reasons like people might not be able to do it at every single visit or it may not be necessary.

For example, if somebody's coming in for an acute visit or simply coming in for education or some other thing, but that reaching thus far would mean that people were doing it the majority of the time.

And one other thing for the committee to consider is that there's a little bit of a circular through argument in the data that we present because we're really only able to test the e-Measure in sites in which disease activity is routinely being performed, but actually our benchmark data in terms of ACR National Surveys show that over 40 percent of rheumatologists don't perform any disease activity assessment.

So that -- you know, their performance would be zero and obviously we can't perform e-Measure testing among those sites. So this is an overestimate. It's clearly an overestimate of national performances in terms of considering the gap.

Anyway, hopefully that helps answer some questions.

Female: OK. Does the committee have any other comments on importance? OK. So we'll move on to scientific acceptability.

And so -- and I made this comment on a couple of things. So one of the statements that bothered me just a little bit, and it's probably because I don't -

- well, it's just the fact that when it -- they were talk -- when we were talking about reliability, it was commonly stated that this -- since it's done electronically that it would be reliable; therefore, go on to the validity testing.

So I'm not sure if that really gets quite at reliability. Other comments? Again, there was a little bit of disagreement. Some reported this as high, some as low as far as the reliability statements. Any committee comments?

(John): Hey, this is (John). I had the same issue that it was an assumption and I just wanted to know if it was -- is that a reasonable assumption that an e-Data poll is considered to be reliable in and of itself.

And secondly that they were actually using a reliability measure, which seems kind of reasonable as a measure of validity that they reported campus scores on comparing a front-end poll from the e-Data poll, and then use that as a reflection of the validity of the process, which I guess, well, strictly when you do reliability tests, it's not a measure of validity in the context that they used. It seems like it would be reasonable.

(Angela): And -- this is (Angela). I just wanted to say that we do -- when we're reviewing the e-Measures, look at the repeatability primarily under reliability. And so we have had the tendency to say that we look mostly at the validity, which is key in the e-Measures, and we can provide additional guidance at the in-person.

Female: OK. On -- so moving on to validity, just generally it looks like there's low rating, moderate rating. One comment that I had is that the validity testing did show differences between the sites, but does this really -- this difference in numbers, does it really tell us how sites actually perform?

And I'm not sure that that's something that's easily measured. So is this kind of the best measure to get at validity?

(Angela): And I might throw that to the developer just to discuss the validity testing.

Female: That would be helpful.

Female: Yes.

Female: Can you -- can you just clarify your question for me and I'll see if I can try to answer it?

Female: Well -- so the information that we have that was provided told us that basically the way I interpret it is that this is a valid measure because we can show that sites actually perform differently, but I'm not certain how that demonstrates exactly how each site performed.

And again, I'm not sure how you would get at that, you know? Is there another measure that actually tells how the -- tells how each of those sites performs and does that measure actually measure what's been reported or actually is it similar to what has been provided in the information to the committee?

Female: So we provide performance by -- at each site at the aggregate level. The information that you don't have, but that we have is individual provider performance within the site. We make some global summaries about how there's significant differences in individual provider performance within each site.

Does that help answer your question?

Female: I think so. I think so.

(Janeuse): So a question over here -- (Janeuse). In the report, you know, the paragraph that you have mentioned under performance measure score of validity, there was mention of one site was instructed to blind reviewers to the results of the automated report.

Does this also capture the validity aspect? Because the -- I mean, integrator with reliability is actual (inaudible) over there.

Female: Yes. So just to be as thorough as possible, we at one of the sites had, you know, complete blinded front-end review of the -- of the EHR, and it was actually surprising how high the agreement was between the electronic data

pools -- automatic data pool, and somebody just reading the clinician note and looking at the EHR.

Female: OK. Shall we move to the -- does anyone have any other things to add for validity? There's some long discussion under 2B, 3 through 7. So I'm afraid I might miss there. Are we OK to move on to feasibility?

(Angela): Yes, if there's...

Female: Yes, I think so.

Female: So a number of the committee members mentioned that the feasibility might be dependent on kind of the current work flow or the particular facility. So that was something to consider. And in general, it was rated low to moderate across the group.

Already clinical recommendation for care, but not necessarily routinely generated. Data field would need to be added to EMRs. And some of the sites indicated that it would take several months to create the work flow changes in order to capture this data, technical challenges.

I think that sums up feasibility there. Anything to add, team?

Female: No.

Male 1: No.

Male 2: Hello?

Female: Yes.

(Drew Aswani): I was calling in for the work group two call. Am I calling the wrong number?

Female: This is correct. Who just joined?

(Drew Aswani): (Drew Aswani). I'm sorry, I couldn't hear anybody else.

Female: Oh. OK. So we -- if you're just joining, we're just in progress of reviewing measure number 2523 and we're at feasibility.

(Drew Aswani): No problem. Thank you. I'm just listening in.

Female: All right. So I don't think there was additional comments on feasibility, so I'll move on down to usability and use. And the -- in general -- so there was a number of planned use that was provided by the developer.

They stated that there was no -- well, this gets to the harm -- the harmonization -- stated that there was no competing -- I'm sorry, I just lost my place -- no competing measure, but they mentioned a current measure used in PQRS. So there's a little discrepancy there.

A couple of comments is their benefit from doing this compared to personnel costs to carrying out this measure, and this is a -- again, referring back to the 50 percent of visits being used as a standard.

I think we discussed that. I don't know if we need to discuss it again under here. And then it should be noted that Medicare will be requiring performance of disease activity assessment for reimbursement in the future potentially.

Anything else from the committee that I may have skipped over? OK. If it -- if there are no additional comments on this one, we can move on. There are no competing measures for this one.

So it did -- but they did mention, just to come back to that last statement that you just made, they did mention a current measure being used in the PQRS program. Yes, it was within the document.

(Angela): Yes, that is correct. There's a -- and I'll throw it to the developer as well. There's a comparable measure that is similar in the PQRS program. Does the developer want to speak to that?

(Marci): I can speak to that. Excuse me. So in the PQRS program, there's a measure that states that these activities should be measured once a year in patients with rheumatoid arthritis and classified as low, moderate or high. In the presence

of those words, low, moderate or high, qualifies as they're passing the measure.

And over the last several years, we, again, through this consensus development process have heard loud and clear from rheumatologists that rather than using a standardized score, which is being proposed in this new measure that actually has specific cup points for what we mean by low, moderate and high, that that older PQRS measure lacks specificity and is open to sort of subjective interpretation.

And so we see this measure as sort of the next generation disease activity measure. It's not a directly competing measure because I think that although the general concept of disease activity measurement is the same, this measure adds a considerable amount of specificity and response to some of the criticisms that we've heard about the previous measure.

Female: And so to add to that, actually, I was part of the PQRS testing at UCLA and I can attest to that that, you know, categorizing it at low, moderate, high didn't really do much. That's why we were using the CDI at each of the visits to add to that.

So this measure definitely adds that specificity, but again, you know, I think where the issue will be is standardizing it across all practitioners and making sure that we are able to capture that data depending on whether it's a, you know, electronic data capture versus, you know, paper.

(Jason Masusik): This is (Jason). And I guess this would be something for the developers, too. I found it interesting that in the questionnaires that went out to the sites that we got to review about whether or not any of these measures would be feasible that when the sites said that they didn't think they would be feasible for measurement or that they didn't think they accurately reflected quality that was intended, to have some further discussion about those points to say, like, yes, OK. So some of the sites felt that this isn't a feasible thing for us to obtain, but here's why we think that it might still be feasible because when you have two out of three sites recording that a measure isn't feasible, I think

it makes it very hard for the standing committee to go ahead and say, yes, it's entirely feasible even though two-thirds of the site say that it's not.

Female: That's an excellent point. Any other good points about 2523? From the committee?

Male 1: No.

Female: OK. So that would move us onto measure number 2524, rheumatoid arthritis functional status assessment, and we have, let's see, other lead discussants, (Kelly Clayton) and secondary, (Jason).

(Kelly Clayton): This is (Kelly). The measure is looking at a percentage of patients 18 or older who have been diagnosed with R.A., and which a functional status assessment was performed at least once during the measurement period.

They were looking at the patient reported outcomes of the measurements, but it was more of the (inaudible), we believe, and then -- so the numerators were the number of patients with functional status assessments documented once during the measurement period.

Functional statuses could be a festering one of a number of valid and reliable instruments. They (inaudible) one particular one. Then that denominator with any patient 18 or older with a diagnose of R.A. who had been seen for 2 or more face-to-face encounters and there were no exclusions listed.

Let's see. Looking at the importance to the measure, we had a variety of comments. There was listed as -- looking at the quality construct, it was high priority, but we felt it was moderate, I guess, evidence.

There was a fault that it was lacking specific information regarding the measure. It -- let's see. The functional statuses were recommended by the ACR, the Canadian and the European guidelines.

The prevention or outcome is, you know, they're trying to prevent the total disability or disability of a patient. The core functional status was associated

with increased mortality and that space elevated with either poor disease prognosis or, I guess, lack of treatment or changes with treatment.

There was no direct -- let's see. Some of the other comments were no direct evidence measuring the FSA improved outcomes. There seems to be information on the literature that patient literacy levels, even they would, you know -- issues could affect it.

Then -- let's see. One had mentioned that in the published guidelines you'd have to perform these every one to three months; during the active disease in every three to six months during remission or low activity.

Let's see. One found it was similar to 2523 looking at 1B. Let's see. Overall, I think kind of looking at this we felt there was moderate evidence that -- I know a lot of the articles I have looked at, they had ranged from, you know -- there wasn't exact data to state that it was prevent total disability, but that they felt that using the practice of these assessments was prevents (inaudible) long-term outcome.

(Jason Masusik): This is (Jason). I'd like to hop in here at this point, if I could. And I think that the evidence that they presented all show that the functional status assessment predicted long-term morbidity, but they never provided any evidence that showed that getting a functional status assessment once a year actually improved the outcome. And that's what I think we're looking at here is that this is not -- this isn't a measure of the -- the patient-related outcome; this is a measure of the process of performing a functional status evaluation.

So I really felt like the evidence was insufficient for this one. I -- I -- I think that moderate would be, you know, if you looked at -- if you looked at what their outcomes -- if you're measuring their -- their disease remission rate and things like that, that -- certainly that does have patient-related outcomes but that's not what they're looking. They're looking at whether or not facilities are -- this goes back to the -- the point, I think, on the initial -- on the algorithm where it's like -- if your target is having control of blood pressures, is it enough to have a process that just says, "Do you measure blood pressures

once a year," you know? Is the measurement enough to actually indicate a quality improvement for the patient?

(John Ventura): This is John . I -- I had the same question.

I -- the -- I guess the forgiving factor for me was that there isn't a downside so much. I mean, they address some of those -- the health literacy and the time and the language under feasibility so I didn't see much of a downside. But I agree with -- with Jason's comment.

Female: Were there additional comments from the rest of the committee members?

So (inaudible), I think we can move onto our next criteria , which would be scientific acceptability.

And apologies to John Ventura. I know that...

(LAUGHTER)

... you are the lead discussant for this measure.

Male: I'm sorry.

(John Ventura): No, that's OK. She -- Kelly can continue on.

Male: OK. When looking at the validity , at least from the information I saw, it kept referring to -- yeah. Go ahead and -- let's see.

Sorry. When looking at the reliability, it kept referring to the validity, which I -- I know some of the other measurers had that same problem.

The -- let's see. Some of the comments from our group were consistent results across the study reflect the focus of a measure, which would be functional -- physical function, which leads to the long-term health outcomes. There were comments that the PROs need to be clearly defined, which I would have to agree with whoever put that comment.

And then there was also a comment that someone was unclear about their assumption that because it's a data pull from the EHR records, the reliability is assumed to be 100 percent.

I know that was also a question of mine. You know, were they just implementing or looking at PROs that were (inaudible) the patient via tablets or iPads, you know? Were there paper ones used that -- there's the assumption that the information could've been mis-typed if it was transitioned to the EHR.

And then someone else had put the -- under the performance measure scoreboard that they -- they do reports on reliability (inaudible) a data extraction for the PRO.

So I don't know if you guys want to jump in here. So it did have a -- the -- the GEPA was the 94.95 . Those 97 points -- 3 percent of agreement between the automated reports in front-end extractions but when they went and reviewed those, the, you know, data was, for the most part, a good match. But then they estimated the missing data to be about 6 percent.

So -- and then it was also noted that there was a variance in performance among sites. I know with the (inaudible) site, they didn't pull information from because it said it had been entered in the wrong field. So...

Female: Comments from the other discussants?

(John Ventura): I -- I guess I -- I just -- I had a question on -- on -- the validity testing was -- was largely done as a consensus process, you know, certainly following rand appropriate in this criteria. And I wasn't quite sure how -- how strongly that that method of validity testing is scored

Male: So -- this is (Angela) .

We do accept face validity testing although it's not the highest testing that we look at but we do accept it.

(John Ventura): OK.

Female: Additional comments about validity testing?

Male: Did -- this is (Angela) again, sorry.

Just want to check with the developers to see if they wanted to comment regarding the testing, especially the 7.6 disagreement and validity score.

Male: Well, I think one thing that might be helpful to clarify for the committee, perhaps before an email, is -- is how some of these testing definitions differ between sort of a standard submission and e-measure submission.

I think we tried. Hopefully we succeeded in following the e-measure testing guidance as -- as -- as closely as possible. And actually the AMA-PCPI were consultants on this since they have more experience and have gone through all these materials to make sure that things are entered in sort of the right place.

So this question about, you know, why is the reliability put into the validity field -- and I think that some of this is -- you know, definitions and the e-measure testing, which is the electronic pool is felt to be reliable since it's -- since it's an -- a computer algorithm that should be reproducible and that doing the front-end versus the -- versus the automated pool extraction is -- is reliability testing. But in -- in -- in terms of thinking about psychometrics, it's also a -- in research, we'd actually refer to this as content validity.

So I think there's a lot of confusion around that. I'm sure if -- if we put things in the wrong place or if, you know, we just need more clarification about where these things should -- should appear.

Male: Hi, this is (Angela) .

(inaudible) -- we do have -- provided that guidance to the committee and it is also just for the committee to note that the validity is entered in the correct location. So I thought that was a question. It is in the correct place.

And what the developer has said about reliability is correct, that we really kind of assume that the repeatability is there for these electronic measures or

e-measures and the focus will be validity testing and the -- I mean, and the validity testing results.

And there was a question about the disagreement found between the automated report in the front-end attraction in -- in 7.6 percent of the testing sample. I was kind of asking a question if there were additional comments the developer could make around that in the validity testing portion.

Male: So I believe that the disagreement was around situations in which the functional status assessment was entered as free text into a progress note and did not appear within the structured feel that populates the automated data pull report. And this, you know, just an -- an issue with the way that EHR are -- are currently structured and the work flows that are in place.

But some sites have perfected their workflows to get the sequence in the right place and -- and others have not. And we can give you more details about that during the in-person meeting.

Female: Great. That sounds great. Thank you. Additional questions about (inaudible)? OK, then it sounds like we can move on to feasibility and our lead discussion.

Female: I was going to say let -- I was going to let one of the guys do it in here.

Female: (John)?

Male: Go ahead, (John).

(John): OK. I feel completely ignorant in that I didn't see where to pick up all the group comments. I saw where they were listed and even graphed out, but I didn't see, as you're showing them now on the screen, but I'll do what I can here and take a look at the screen.

In terms of the feasibility chart that they showed, the biggest issue that came up was relative to -- the biggest issue that came up was relative to the time of implementation that they felt -- at least two of the three sites felt that it was feasible to do the implementation. But they felt that it was going to take at

least several weeks to overcome some of the technical aspects and probably several months to get the personnel training.

They did state that they had addressed, using specific patient reported outcome, health literacy, time of impact, other languages and ease of scoring, so that they did take those factors into the account.

(Jason Masusik): Yeah, this is (Jason). I -- yeah -- no, I agree with those sentiments. I think that the feasibility results are telling. They do the study -- the survey. Is the measure feasible for you to implement today? And if no, it's a main challenge, you know, that they -- the sites agree that there are some significant challenges to implementation which I think is fine as long as we hear how it's going to, all of a sudden, become feasible for these organizations to do that. I don't have any problem as long as there's some type of suitable explanation or something else in place that tells us why.

Also I think that the flexibility that the writers of this measure allow the individual cites to choose which functional status assessment they want to use is great. But I think that when we're assessing some of this and you're talking about feasibility, you have to take into account that some of the measures are a little bit more cumbersome, some of them could take up to five minutes for the people to complete. The clinicians -- actually, some of the measures require clinician training to be able to score them appropriately. So it does impact the feasibility a little bit. I still think it is a feasible measure. It could be a feasible measure. But I'd like to hear more about how some of those things could be overcome for the sites.

Female: And could we hear from a developer a little bit about that?

Female: So this is (inaudible) again. I think in the last several years, its people have been getting their EHRs up and running and just doing basic functions. Some of these ideas of, for example, capturing patient -- reported outcomes of this for and participating in a national registry that can then aggregate outcomes and provide benchmarking information to physicians visiting more far out. But I think that things are rapidly changing. And what we're seeing is that some EHR vendors, and I'll just give an example of Epic, are incorporating

some of these PROs into their product. And you know, as that increases, I think the accessibility of some of these things is going to change. And so, you know, one questions -- this is a service theoretical question is whether we sort of want measures that set a really low bar and sort of reflect the current reality of what EHRs look like or if we want to get to the meaningful part of meaningful use and actually create things that are meaningful which I think is moving more towards outcomes.

And so I think all of the plans about current feasibility that were brought up in the feasibility testing are true. We don't want to minimize them, because there are limitations. There's workforce training issues there's sort of local implementation issues. Those are all obstacles. But we're hoping to sort of provide a vision and also examples, because we have examples of people successfully implementing these workflows and are using these things in (inaudible) practice today and what that might look like in the coming years.

So we can provide more information as well at (inaudible).

Male: Yes. I think that's good. But the other thing that you just said, and I want to come back to that real quick is that you said that you will want to track outcomes. And yet, all of -- several of these measures now are not tracking any type of outcome. These are all process measurements. And in this particular case, we're not tracking the outcome of disease remission. You're looking to track whether or not people actually perform a functional status assessment. And I want to be clear about the distinction between the two. Because the one is certainly -- is a very, very high priority issue, tracking the outcome. The other -- the process measurements, when there is a way to possibly assess the outcome, you know, should be a secondary thing that we shouldn't necessarily be as focused on on the process measure, if we can find some way to actually get a good patient-related outcome.

Female: So I guess, just to make a comment on that. One of the things that became clear through this measure development process which was lengthy and we tried to make it as comprehensive as possible, is that there's essentially been no work in case mix adjustment or other methods to be able to get to a outcome performance measure. And one of the obstacles is that people are

not actually measuring outcomes routinely in clinical practice. So although I understand that that's an important future thing to get to, there's no way. I mean, we've I think turned over every stone to try to see if there's any we could come up with an outcome measure and there just is not -- there isn't data and there's -- certainly were not resources to perform that kind of sort of de novo risk adjustment within the confines of this process.

And so, I do think that one of the things that this measure is moving towards is standard collection of outcomes. And if you look at the e-measure specifications, it's not a yes-no, checkbox. It's actually that a score was entered. And you can imagine that in the context of a registry, if you have a score, then things like doing some of that risk adjustment and feeding that information back becomes possible. So anyway, just that's sort of what we were thinking in designing this step in measure development.

Female: Thank you. Any other comments about feasibility? With that, we can move on to use and usability.

Male: And the issues that came up regarding usability is that it's anticipated that this e-measure will be incorporated into a national registry and they -- and it was the same issue as the prior measure that we looked at that there is a PQRS measure. But this is felt to be an enhancement that adds specificity. So that it will be a -- so that it does meet the criteria for usability. And I don't know if anyone else has any other comments.

Male: I think one of the comments that was in there that I agree with is the fact that I don't know it's going to -- I don't know if it really will alter care.

Male: Oh.

Male: You know. And you know the one thing I worry about is, you know, measuring things that really don't have impact. And I think you can use it. But the question is on a larger scale, will it really make a difference? Just throwing it out there.

Female: Additional comments about usability and use?

OK. (Just) time. If there are other comments, please let us know. And we have collected quite a few questions that we'll be circulating amongst ourselves at the committee, as well as the developer, and prior to the in-person meeting.

So, this moves us to our public comment portion of the agenda. And, Operator...

Operator: The call lines are open.

Female: ... are the lines open? Yes.

Operator: Yes, ma'am, all lines are open.

Female: Thank you.

(Katie): Any public comments?

OK. So, this is (Katie). Thank you all for attending today's call and for all the preparation -- your preliminary evaluation survey. Hopefully, you have been receiving information regarding booking your travel, the hotels and flights.

If you have not, just send us an e-mail, and we'll make sure that our meeting -- meetings department follows up with you.

Other than that, please send us any other questions you have. And we look forward to seeing you in a couple of weeks at our meeting.

Male: Thank you.

Male: Thank you very much.

(Katie): OK. Thanks you so much.

Thanks. Bye.

Female: (Katie), this is (Pujah) here.

(Katie): Yes?

(Huja Kana): I haven't received anything regarding the travel. So, if you could forward that information for me, please?

(Katie): OK. We certainly will. And just as an FYI, it'll come from a meetings@qualityforums e-mail address.

(Huja Kana): OK. OK, sounds good.

(Katie): But we'll follow up on that for you.

(Huja Kana): All right, thank you so much.

John: Katie, this is John. I did get the information regarding flights, but I haven't received anything regarding a hotel, although I didn't check my e-mails today.

(Katie): OK.

(Huja Kana): Well, when did this go out, by the way?

(Katie): I'm not sure.

Female: Last week.

(Katie): I think last week sometime. But we'll -- we'll definitely follow up with our meetings department and try to find out more information on when this was sent out, and...

(Huja Kana): OK.

(Katie): Also, maybe want to check your junk mail, just in case it...

(Huja Kana): Yeah.

(Katie): ... was sent there...

John: Oh.

(Katie): ... or your spam mail.

John: OK.

(Huja Kana): OK.

John: Thank you very much.

(Katie): Sure. All right. Thank you.

END