NATIONAL QUALITY FORUM

Moderator: Measure Developer Maintenance
April 6, 2017
12:00 p.m. ET

Operator: This is conference #: 82005464.

Operator: Welcome, everyone. The webcast is about to begin. Please note today's call is being recorded. Please stand by.

(Katy Streeter): Hello. Good afternoon, everyone. This is (Katy Streeter), senior project manager at NQF. You are attending the Musculoskeletal Standing Committee Post-Comment Call to Discuss Public and Member Comments.

The purpose of today's call is to review and discuss comments received during the post-evaluation public-to-member comment period. We'll be asking the committee to provide input on responses to the post-evaluation comments, re-voting on the scientific acceptability criteria for measure 0052 and determine whether to reconsider measure number 0514 as requested by the measure steward. And, of course, we'll take a deeper dive into all of that in just a few moments.

With me here in the room is Karen Johnson, senior director at NQF. And now we'll just take a quick roll of the committee to see who is on the line with us starting with our co-chair, Roger Chou.

Roger Chou: Yes, I'm here.

(Katy Streeter): Kim Templeton?

(Off-Mic)

Marcie Harris Hayes: … on the speakers. Will I be able to hear it through the phone?

(Katy Streeter): Hi, Kim. Do you think that was Kim? Kim, was that you?

Marcie Harris Hayes: Marcie just joined if that's.

(Katy Streeter): OK. Hi, Marcie.

Marcie Harris Hayes: Hi.

(Katy Streeter): Thiru Annaswamy?

Thiru Annaswamy: I'm here.

(Katy Streeter): OK. Carlos Bagley? Steve Brotman?

Steve Brotman: I'm here.

(Katy Streeter): Sean Bryan? Craig Butler? Kelly Clayton? James Daniels? Christian Dodge?

Christian Dodge: I'm here.

(Katy Streeter): Katherine Gray?

Katherine Gray: Here.

(Katy Streeter): We know Marcie is with us. Mark Jarrett? Puja Khanna? Wendy Marinkovich?

Wendy Marinkovich: I'm on.

(Katy Streeter): Jason Matuszak?

Jason Matuszak: Yes. I'm here.

(Katy Streeter): Art Schuna?

Art Schuna: Here.

(Katy Streeter): John Ventura?

John Ventura: I'm here.

(Katy Streeter): Christopher Visco? Thanks. Did anyone else join whose name I did not call?

OK. And just to remind you, this call is open to the public and we also have several measure developers available that will be participating today and will be able to answer any questions as we move along.

Before we move on to beginning our discussion about measure 0052, use of imaging for low back pain, it is a measure from NCQA – I'll pause and see if our co-chairs are – I guess, Roger has any opening remarks.

Roger Chou: No, not really. Just thanks, everybody, for being on the call again and helping us to work through these measures again. There were some public comments, about 50/50 I think that – about 50 percent thought that they agreed with the decision not to move forward with the endorsement when we last looked at the measures and the other 50 percent wanted us to endorse the measures and I believe that the measure developers have responded to the public comments are well which we'll be talking about. So, anyway, thanks everybody for participating.

(Katy Streeter): Thanks, Roger. So, now, we'll begin our discussion about measure 0052 and I'm going to turn it over to Karen Johnson who will walk us through that.

Karen Johnson: Thank you, (Katy). And before we get into a committee discussion, any expert discussion, I want to just remind everybody what happened back in December regarding this measure. It is included in a little bit more abbreviated form in your memo, but if you recall back in December, you voted – the vote was 12 insufficient for validity and 5 low and then 17 insufficient for reliability. And, again, those are must-pass criteria, so the measure at that point went down.

We, however, think that many of the – maybe perhaps the majority of the insufficient votes for the validity sub-criteria, this may have been because of our NQF staff preliminary rating of insufficient.

And after looking again at the submission, we actually changed our mind and we mentioned this to you before and it is in your memo. But we no longer consider the data element validity testing presented by the developer to be insufficient.

So let me just walk you through our thinking just a little bit there so that everybody is on the same page. The developer did make for …

(Off-Mic)

Karen Johnson: … NQF …

(Off-Mic)

Karen Johnson: … material changes to the measure when they added additional exclusions. And if there is what we call a material change, then we do require updated testing. That's the part of our criteria that has always been around.

The information that was presented in the testing section of the form was actually not updated to reflect the changes in the measure. However, in – a little bit further down in the form, still under the validity section, the developer provided information about some of the additional exclusions that they did add in. And those additional information – the data that they provided does actually address data element validation.

So, in our preliminary analysis specifically, we had noted that, you know, we're look – you know, we are looking for sensitivity and specificity statistics. Now this kind of statistics help us answer questions like to what extent are the diagnoses that are recorded in the medical record also reflected in the claims and kind of the flip towards – or the diagnoses that are reported in claims not actually present in the medical record.

So even though the developer didn't actually provide, you know, formal sensitivity specificity statistics, they did provide enough information to shed light on those questions and you'll find that information, again, just for your reference, in their table seven. So, again, we would like you to disregard our previous guidance and our previous rating of insufficient for validity and take into account our current rating which would be moderate.

So what does that mean for you in terms of the re-vote today? Well, first, we are specifically asking you to re-vote because of our change in guidance. And so, a couple of things to think about when you're re-voting.

First, that it doesn't mean that you cannot still vote insufficient. We think that the data that are provided should (rely) on how well the excluded conditions are represented in claims data. However, there may be other reasons if you might want to consider the information to be insufficient.

For example, the data are old even – you know, and you might really want to see updated or more current data even though this is not a requirements of NQF. But this also doesn't mean that you have to rate the measure validity as moderate. So, again, just look at the data that are provided and both based on those results.

So just a couple extra things. First of all, NCQA, after the call let us know that they were going to try to get more updated information and we actually discussed that very briefly in the December call. And they have, in fact, provided some information that may help you with your vote and we can look at that in a little bit more detail later.

And then, finally – and this is very (meaty), but I need to go ahead and explain it. We're going to – like we did in the December call, we are going to talk about validity first because that's where the data really are that we need you to look at.

The reliability testing data that – or in the current submission is actually still insufficient because the results there don't reflect the updated specifications. However, and this is the goal of NQF guidance, you can actually use the data element validity testing results to decide whether or not you are OK with

passing the measure on reliability. So when – if it passes the validity when you vote, then we'll go on to reliability, so again, kind of flipping it.

When you vote on reliability, you'll be thinking about the validity testing results for the data element, as well as whether or not the specifications are clear and precise enough that the measure can be consistent. So I know that's a lot and it's kind of, you know, in the – in the (set weeds). But let me stop and see if there's any questions on why we reversed our guidance and what we're going to ask you to do.

Thiru Annaswamy: This is Thiru Annaswamy. Can you quickly walk us through the table seven you alluded to?

Karen Johnson: Sure. (Katy), can you find that in the …

(Off-Mic)

Karen Johnson: And (Katy) is bringing that up on the screen. And I don't have the page number. She might be able to tell you the page number.

But, basically, table seven, the title of the table is "Source of Red Flag Exclusion Diagnosis Among LBP Patients." And the – they have several of the elements that are used as exclusions included in this table and in – I think in red, if I'm correct, OK, they actually added in to this table that they had originally submitted – they added in the new stuff that are – that are new exclusions.

And, basically, what it is telling you – what this table is telling you is that if you look only at claims data, you can see that, for example, prior cancer, I think it's 0.9 – sorry, just a second. Let me orient myself to this table. Apologies. My table is going over two pages because I'm getting a lot.

You know, what this is telling you is it's telling you a couple of things. It's giving you a flavor of how frequent the exclusions are and it's showing you where you would pick up those exclusions, so, in other words, if you would pick them up only in administrative data or only in the medical record or if they were in both. And what we would really like to see is either they are in

both the administrative and medical records or in neither of them, so, in other words, the two data sources are great.

So, for example, the first one on the list there, recent trauma, what you're seeing is, is that for 80 percent of the records – and NCQA folks, you can tell me if I'm misinterpreting this table – but it looks like 80 percent of the time recent trauma is in – not included in the either of the claims data or the administrative record and in 0.2 percent of the time it's included in – it is included …

(Off-Mic)

Karen Johnson: But 90 percent of the time it is – it's only found in the medical records. It's not found in claims. And that's how it works all the way down with the various exclusions.

Let me stop there and see – folks from NCQA, did I interpret that table correctly? And, if not, apologies and would you fix it for us.

(Jenna): Karen, this is (Jenna) from NCQA. You interpreted the table correctly. Thanks.

Karen Johnson: Thank you.

Roger Chou: Yes. This is – sorry, this is Roger. Just wanted to – you know, we have looked at this table I think back in December or whatever it was. And I think many of us were struck by – you know, I think NCQA has said, you know, that there aren't a lot of red flags, you know, that you would be missing by just looking at the administrative data.

And a number of us looked at this table and said, well, 19 percent of recent trauma – first of all, that's a weird number. That seems excessively high. But if it's accurate and we haven't seen any explanation otherwise, but if it's accurate, that's a pretty high proportion of red flags that aren't picked up in administrative data. So this was a concern from the group back when we discussed it before and I don't see that that's been parsed down any further.

Karen Johnson: And, (Jenna), would you like to comment on that? I know the last time we had talked about your feedback you would have from users of the measure, but I don't think the newest data was able to speak to that particular piece, right?

(Jenna): That's right. The data we have does not go to that level of detail, so we're not able to parse that out. I mean, I think the two things that we can say is that this testing data is fairly old from back in 2014 and we would expect that claims data would have gotten better since there. There are ICD-9, ICD-10 codes that can be used for this.

And the second thing is that we are not getting a lot of feedback through our online system through which users who have been measured can submit feedback. We aren't getting a lot of concerns through that, that they're not able to capture trauma. So, fortunately, that's the most I can say about it.

Thiru Annaswamy: Another question from Thiru Annaswamy. Would you mind going over how you determine this to be a moderate validity as opposed to insufficient on the December webinar?

Karen Johnson: Sure. So, as I said before, really what we're interested in is kind of the congruence between what's found in medical records and – versus claims for this kind of measure. So even though they didn't do updated testing per se and give us those formal statistics, the information that is found on this table seven really does tell you what was, you know, found in either both or only one of the two data sources.

So it does everything except actually compute those statistics for you and it does it specifically for the new exclusions that were added to the measure which is the ones that we would, you know, really be interested in because that's the new stuff with the measure.

So in terms of why moderate and not high, a couple of reasons. This is what we would consider data element information and that, you know, the highest you could rate it would be moderate because it's only data element information.

You know, the 19.3, you know, we could have, you know, teed on the 19.3 and perhaps not done moderate. But I think we were swayed by (Jenna)'s discussion of their – the feedback that they have from users. You know, I don't have right in front of me – and (Katy) maybe you can find it – I don't remember how large they attest – it's in the 431 record. Yes, it could have been just bad luck in the draw potentially that something was going wrong with the trauma.

We don't know, but after several years of use of this data, they're not getting complaints about that particular thing. So that swayed after preliminary analysis to go moderate as opposed to lower insufficient.

Thiru Annaswamy: Thank you.

Karen Johnson: Does anybody else have any questions about, you know, what we discussed in terms of the – you know, really the NQF reversal of our guidance? We do want to be very transparent. When we make a mistake we wanted to own up to that and make sure everybody understands what we were thinking and we want to make sure that, you know, folks aren't penalized because we made a mistake, so.

OK. Hearing no other questions, I think we are definitely at – we're planning on asking you to vote a re-vote. However, (Katy) just informed me that we only right now have 12 people – voting members on our – on our call and you may recall that we need a quorum of 13.

So let me pause just a minute and, (Katy), do you want to recount? (Shawnn), do you want to help us recount and make sure we haven't missed somebody that give us 13?

(Shawnn Bittorie): Sure, Karen. At the moment, I'm still showing 13 and I'm just going to rattle off the names of the folks that were missing just in case somebody's dialed in and maybe not logged in.

So we're missing Carlos Bagley, Catherine Roberts, Christopher Visco, James Daniels, Kelly Clayton, Linda Davis, Mark Jarrett, Puja Khanna and Sean Bryan.

I'm not hearing any responses on those. If I've said your name and you're not dialed in, you're just logged into the conference maybe under a different ID, if you could send a message through the chat area. Otherwise, we are at 12. And thank you, Craig, we do see your message and I did have you already in the count.

Karen Johnson: OK. Well, that's unfortunate. Thank you, (Shawnn). It is unfortunate. What that means is we can't actually do the vote today on the call. We are required to have quorum. The quorum has to be more than – remind me, (Katy) – more than 66 percent, the seated members? So 13 was our magic number.

So what does that mean in terms of your vote? Well, that means that we get a reprieve. You get to vote in the next day or two. I think, (Katy), do you have SurveyMonkey ready to send out or close to it?

(Katy Streeter): Yes, probably Monday because we'd like to have the transcript.

Karen Johnson: Yes.

(Katy Streeter): Yes.

Thiru Annaswamy: You can invoke the nuclear option like the Senate?

Karen Johnson: We might like to, yes. What we would like – what we will do is probably send out the survey to all of you. It will be a SurveyMonkey survey. It will come from (Katy) or from the musculoskeletal mailbox.

And along with that survey will be a copy of the transcript and this memo. So we will be sending it out to all members of the committee. Hopefully, all members of the committee will be able to read the transcript and read the memo and provide a vote. But we won't do that on the call today.

However, on the call today, we would like those of you who are on the call to talk just a little bit about your thinking in terms of the validity information that is available. Do you have any questions, any concerns about validity?

Do you want to discuss – and actually, we still need to go back to (Jenna) and let her talk about – or apologies, (Jenna), that we kind of got off track a little bit – kind of briefly introduce what you provided to the committee and any discussion that you want to have and that you want your fellow committee members to react to when they read the transcript. So I'll just open it up for discussion about validity.

Roger Chou: This is Roger. Don't we have some committee members kind of assigned to walk us through? Maybe we should just have them walk us through their assessments and then we can have a discussion.

Karen Johnson: Perfect. (Katy), do you remember? Is it – who is going to be the lead discussant …

(Off-Mic)

(Katy Streeter): We have – the re-discussants Chris Visco and Carlos Bagley and Cath Roberts, but she was not able to join the call today.

Roger Chou: Yes. I think none of those three are on the call.

Karen Johnson: That was unexpected. We knew Cath was not going to make it. OK. Would anybody else like to discuss their thoughts?

Thiru Annaswamy: Some of my questions about the table seven was essentially try to clarify the issue. This is Thiru Annaswamy again. So those were my questions. I don't have any other pressing questions at this time.

Karen Johnson: OK.

Katherine Gray: This is Katherine Gray. I just was wondering, the table seven, had they not submitted any information that – would that have caused you to re-evaluate your other decision – your insufficient or was it the new information that changed your opinion?

Karen Johnson: It was kind of two things. So the stuff that's in red, if they hadn't put that stuff in there then it would have been insufficient, period. But then, stuff in

red does talk – you know, it speaks to the additional things that they added to the measure.

So we had seen that and, to be honest with you, I was just being a little too geeky I think in really wanting to see the sensitivity specificity, you know, formal statistics. But, yes, what they gave us I think is enough to answer what sensitivity and specificity is trying to tell us. So it was really my mistake in being a little bit too hard on the statistics that were provided. Does that make sense?

Katherine Gray: Yes. And I mean, I don't know if you actually calculated it or not, but did you come up with some overall sensitivity and specificity then given the information in the table?

Karen Johnson: No, I didn't. We had kind of hope that the NCQA folks would be able to go back in their files and find the actual (cell sizes). You can think of the number set would go in here as kind of two by two tables. And if they had that data they could – they could have calculated those numbers for us. But I don't know if they were able to go back and find them or maybe they weren't able to. (Jenna), do you know if you guys?

(Jenna): I don't think we have any more data than what's presented here in the table. So I don't think that we have the data needed to get to that level.

Karen Johnson: Yes. I mean, if they had had the actual kind of underlying data I think they could have just – you know, it's a pretty simple calculation. But again, what they have here really does tell the story anyway or pretty close to it I think.

Male: So …

Katherine Gray: And just what – go ahead. Sorry. Minor question. I was just going to ask if you knew – if the N was 431 before were in the blue letters or numbers and the red is the new stuff that was data that was available on that same population of 431 or did the N get larger?

Karen Johnson: (Jenna)?

(Jenna): No. I think basically what had happened was when we originally submitted this, we submitted data on the exclusions that were in the measure, so the total number of cases and things did not change. All we did was we included rows this time for the exclusions that we actually did end up including and others that have been considered at the time the measure was originally tested. So all of – the description of the testing and everything remains the same. We just added some rows here.

Roger Chou: So this is Roger. I mean, I'm still troubled by this very pretty high proportion of people with trauma that are only captured in the medical record and the fact that we still don't have an explanation for it. I don't – I don't think just saying that people haven't complained about it is a – is a very compelling explanation.

And the neurologic impairment was the other factor that we were worried about. You know, you have a four times higher rate of neurologic impairment when you look at the medical record than when you look at the administrative data and there's also been no attempt to parse that out.

We had some questions earlier about what was meant by neurologic impairment because a lot of people will only have radicular leg pain without necessarily impairment and whether that was even included or not in this data and I don't think that's been answered either. So, you know, I still have some questions that I don't really feel have been addressed.

I mean, these are the same things that have come up repeatedly. And again, this is very old data. I think this is from 2002. And so, I also question why there hasn't been any attempt to look at that more recently.

Thiru Annaswamy: Yes. This is Thiru. I echo Roger's comments and also if the – some of the comments earlier on the call today said that there is a plan for additional data and testing to be – to be done. So perhaps it would be more prudent to wait for the new testing updated information and then come back to the committee so we don't have to keep doing this and everybody gets frustrated in the – in the process.

Karen Johnson: So this is Karen and apologies. Maybe I was unclear. Back in December, NCQA had mentioned that they had hoped that they would be able to bring you some new data and they have brought you some new stuff today.

It just doesn't – they weren't able to update table seven. They gave you some additional things that we can still talk about. But at this point, I don't think – well, actually, let me stop there. (Jenna), you'll be able to get new data and run reliability testing data later in the year, but that wouldn't speak to validity would it? I mean …

(Jenna): Right. I mean, what we can – what we're able to do when the plans submit data with these – this specification has been revised that comes to us in June as we would be able to run reliability on those. If we had a similar enough measure where we would expect there to be some sort of – some sort of correlation in rates and performance, we could do an analysis comparing those – comparing those rates as well. This is a hard measure to find a corollary measure to because it is an overused measure. It's about imaging and we only have – we have a fairly small number of overused measures in our set.

Actually, getting down to the level of data element validity, the reason why we don't have newer data on that is that it's a very expensive type of testing to do, you know, to actually go to plans and ask them to identify medical records and then compare it against claims. So that is not something that we are – that we do on a really regular basis.

As far as the getting feedback from plans, this is a really robust system by – through which plans who report are measured. And we have hundreds of plans who are reporting our measures, are able to provide us with real-time feedback and ask questions about our measures. And so, when I say we haven't received a lot of push back, we're not getting questions from plans, we're not hearing complaints from them about this particular exclusion in this measure.

And we do have other measures where we do get feedback. And there are things that are difficult for the plans to do and we hear that feedback, but this

is one where we don't. We also posted the measure for public comment as part of the revision process.

And again, we did not hear from plans saying that they had problems with this exclusion or really the other ones either. So it means that they do take that opportunity to provide feedback when it – there are parts in measures that are real pain points for them.

Bob Rehm: Hi. This is Bob Rehm. I'm with (Jenna) on a variety of measures. And just as a – just to give you some scale, we recently had our 2017 – or it's what we call HEDIS 2018 public comment period.

We received 1,300 public comments. And that was on about 12, 13 different measures, some of them new, some of them re-evaluations. So in our policy clarification support system, which is the online portal that (Jenna) mentioned, we received – I think that's over 10,000 queries a year questioning things from an ICD-10 code, across walk to something else, an interpretation of guidance, interpretation of specification or kind of a like a policy questions at large about the relative merits of a pay per measure.

So this is – this is a fairly robust system that helps generate it. When people are being evaluated, health measure being evaluated why payers for pay per performance, they are very aggressive about saying that's not fair. That measure isn't doing what you said it was and we're going to tell you that you're missing the bow here and we have not heard that on this measure.

Jason Matuszak: This is Jason. I'm just going to interrupt for half a second. I got a hold of Sean Bryan who said he'd be trying to call in momentarily if that helps us get up to quorum.

Sean Bryan: I'm now on the line.

Karen Johnson: Thanks, Sean. Sean, how – did you just now come on or have you been able to follow the discussion?

Sean Bryan: I just got on two minutes ago.

Karen Johnson: OK. We're willing to go ahead and start the – and do the vote online now that we have the 13. But we want to make sure that Sean understands all the discussion. So let's keep going and then we can go back and make sure that Sean gets anything that he – that happened before he came in.

(Shawnn Bittorie): And Karen, if I may. Sean, will you be logging in or will you just be staying on the phone?

Sean Bryan: Yes. Unfortunately, I don't have access to a laptop or desktop computer at the moment so I'll just be on the phone.

(Shawnn Bittorie): OK. So, Karen and (Katy), we can do it a couple of ways. We can – if he's – if Sean is comfortable speaking his vote, we can do it that way or I can have the operator pull him so he can give his vote information to her privately.

Roger Chou: So this is Roger. And thanks for the explanation of the feedback process which I think gives me a little bit of reassurance. But it's still – it's still troublesome that you have actual data that shows that nearly 20 percent of people have a red flag condition that's missed.

You know, you have actual data there. I mean, it's your own data. It's not – I mean, it's not coming from anybody else. Then, your own data shows that a lot of red flag conditions are missed. And then, to say that, well, the plan say it's not a big deal or they haven't complained about it I guess is what you're saying.

You know, that – I think there – it's still problematic. I mean, you know – and, yes, maybe the testing is expensive, but this 15-year-old data. So, you know, I just feel that there should be – that there should have been some attempt at least to parse this out or at least to re-look at what it meant with what – you know, why are there so many trauma cases that were missed.

And I don't think that 5 percent is actually trivial or 4 percent for neurologic impairment either. So, you know, the – it does feel to me that there is still an issue there and it's hard to explain it away by just saying that people haven't complained about it.

(Jenna): So, Roger, let me say a couple more things to try and address this. So one is actually, when we looked at the testing with the new specification that actually we had shortened – if you remember one of the things we did is we shortened the trauma exclusion time period. It used to be a year and we felt like that does not provide enough temporality to the actual back pain, so we shortened it to three months.

And we did see that the number of exclusions actually decreased when that happened which does show – here when it looks as though, for whatever reason, the plans have participated in our testing had a difficult time finding trauma in their claims. But that does not appear to still be the case when we looked at our latest – the plans that participated in our latest – we won't call it testing, but they provided some data to us.

Because if they still – if they were still unable to find trauma in the claims, then shortening that time period would have had no impact on their exclusions and it actually did. So I think that – I mean, I know it's not satisfying, but that – unfortunately, right now, that's the most information I have. I can just – it appears as if whatever was happening in 2002-2003 is not still the case.

The other thing is, yes, while there is a 3 percentage point difference in the rates for admin only and the medical record only for neurologic impairment, these are health plan level measures, so we actually consider exclusions that have less than something like 2.5 percent impact on the final rate to actually be fairly minor because we are talking about fairly large numbers here. This isn't a physician-level measure. So that's just something else to keep in mind.

And for the neurologic impairment, the last thing I'll say is that we are – this is for members who have a diagnosis of cauda equina in the measure as specified now. It is not – it's not just radicular pain or something like that.

Robert Chou: So can you – can you clarify that last point? So does that mean that radicular pain is not considered an exclusion?

(Jenna): Right. When we talk to our experts, we did talk about whether radicular pain on its own should be – is an indication for imaging in the first four weeks. And because radicular pain can be – have an impact on function or it cannot

and there's such a wide variance in it and even the guidelines that we were looking at did not recommend imaging in the first four weeks for radicular symptoms on their own unless they were progressive and getting worse, then we – our expert work group suggested that we should not exclude members just because they have radicular pain.

Robert Chou: OK. And do you have any of the – I don't see the data that you said you just got recently. Is that in the document anywhere or that's not – that's not – wasn't included here?

Karen Johnson: This is Karen. And it is included. (Katy), can you show us the new information that NCQA provided? And apologies, (Jenna). Maybe you can take this point in time and explain once (Katy) brings that up on the screen. You had data from a couple of different health plans that were able to give you some new information?

(Jenna): Sure. So what we were – what we did was we asked a couple of plans – we asked around for a few plans who would be willing to basically share their data with us for their measure submission that, as we said, normally comes in June to see if anyone had already run that data and would be willing to share it with us so we can see how the revised measure specification had an impact on their performance.

So the two-plan tier that's submitted, what – the first table shows how their 2015 rate compared to their 2016. So the 2015 was using the old version of the specs. 2016 is using the revised version of the specs.

And then, actually, if you go to table two, that might be of most interest to the steering committee because here what they did is they actually – since we're looking at plan-level data, it does – the members who are included in the eligible population changes year to year. So what they were actually able to do was they took their 2016 data and just ran the 2015 version of the spec and the 2016 version of the spec.

So here, as you can see, the eligible population did get a little larger and that was because we did add some other ways for members to get into the denominator. We included patients who had a Telehealth visit or physical

therapy with low back – low back pain diagnosis to the denominator. So it wasn't surprising to us that that got larger.

We reduced the timeframe for the trauma exclusion, as I mentioned, and only added a few exclusions to the measure as well. But these were ones we already knew were fairly rare. It was things like major organ transplant, history of prolonged corticosteroid use, HIV.

And so, we sort of knew that that wouldn't have much of an impact on the number of exclusions. And actually, as you see, the number of exclusions decreases and that's because, again, that 12-month exclusion period for the trauma we took down to three months, so again, an expected change.

You see the change in the number of numerator events and increased slightly because the denominator side increased. But then, what's really interesting is that the performance rate ends up staying pretty much the same despite those changes we made to the denominator and exclusions.

Robert Chou: So this is helpful just to look at the number – you know, that you have fewer exclusions because you've shortened, you know, what's excludable in terms of trauma. But it doesn't get at the prior question which is, are you missing trauma by looking at only administrative data, correct?

(Jenna): Correct. I mean, we don't get the level – right. We don't know if there were additional trauma exclusions that could have been in the measure. But we – I mean, you can – you can see that they're able – that they are reporting some using claims now whereas that testing data made it appear that they couldn't find any trauma cases in their claims data.

John Ventura: This is John. Could I ask a question of NCQA relative to what definition of trauma they used and how sensitive were they in defining trauma to do the review of the medical records and could that have accounted? I mean, would've a minor trauma which by most guideline recommendations would not advocate for imaging still fall into that category of being trauma from the medical record review? So, technically, it wouldn't be an exclusion even though it was constituted trauma.

(Jenna): Yes. I would have to go back to the original testing report to see how they defined it. I don't know that off the top of my head unfortunately.

Robert Chou: I think we asked this before. And, you know, that was one of the issues that it wasn't really clear how trauma was defined, so we don't even really know what it means. I think this is one of the questions we've had all along. And again, I think it would be helpful to parse that out a little bit.

And I take the point about neurologic impairment, you know, the overall percentage not being that high. But, I mean, again, the data show that you only captured 25 percent of the people who have neurologic impairment. I mean, that's – you know, that's a – that should be a relatively straightforward red flag to capture. And the fact that administrative data only gets a quarter of those patients, you know, that's – I don't know, that's troubling to me, you know, even though that the overall numbers are not that big.

Anyway, I want to see if other committee members have other comments about the validity. I don't – I don't know if we've talked about reliability yet, but let's maybe finish up with validity to see if there are other comments or questions. And then, I think we were going to have to – I don't know if the next step is to take a vote now that we have 13 people or what.

I thought we usually did reliability first or maybe it's the other way around, I can't remember. But NQF folks, let me know if we – if the order – you know, the procedural stuff. But let me just pause and see if other people have comments about the validity stuff.

John Ventura: Was their face validity testing of 0052?

(Jenna): So – I'm sorry.

Robert Chou: Go ahead. I was just – I think the face validity is based on kind of the match between the specifications and the guidelines and then they also used peer review input and other things, but you guys can talk some more about that.

(Jenna): Yes. We have a number of panels that we review our measures that helped us make sure that we're really specifying what we intend to specify and then also

to assess the clinical appropriateness of the specification. We have a measure-specific panel who has expertise in the area, so we have a musculoskeletal measurement advisory panel who informed the original creation of this measure and then the revisions that it underwent last time.

We also have a larger committee on performance measurement that is a multi-stakeholder group who reviews our measures and approves their use in our programs. We have a technical panel who provides that feedback on whether the specification is really doing what we intend. We have a pharmacy panel and a coding panel so – and a lab panel, so all of those provide input.

And then, we do have, as I mentioned, a public comment period where – I mean, as Bob pointed out, we get quite a lot of traction with our public comment. So those are the – those are the ways in which we do get feedback on these measures and we access the face validity.

John Ventura: Thank you.

(Katy Streeter): And looks like we have a question from Kim. Kim, are you able to speak and ask or would you prefer that I read it off? OK. She may just be on the webinar portion. But Kim did ask a question. Is NCQA going to continue to include Telehealth visits in the denominator? She knows that is concerning as part of evaluating low back pain as a physical exam.

(Jenna): There's a reason why we added that Telehealth. It's not because we thought necessarily that it's an appropriate way to diagnose low back pain, but that any of the encounters for low back pain in this measure are really a proxy for when symptoms start.

We also recognize that patients may have back pain for a period of years and the episode – you know, it's episodic, so they'll have an episode and then it will go away and come back sometime later. So we wanted to basically be as flexible as possible with, first, the identification of low back pain for the denominator because we have heard some feedback that we weren't being flexible enough.

So that's really – that was the driver for including the Telehealth and also the physical therapy. It's not necessarily places where we expect it's – a very initial diagnosis to be made but more that it's a way for – if a patient is scheduling a Telehealth visit, their first way of saying, "Hey, I have low back pain." And so, we wanted to be able to get as close to when the symptoms start as possible.

Karen Johnson: So thank you, (Jenna). This is Karen again. Back to the question about how we usually do things, because the developer was not able to update the reliability testing and do score-level reliability testing, in December, we went ahead and flipped the order that we normally do.

So we talked about validity first because you're going to have to basically take what you think about the validity testing and apply it as you vote on reliability. So that's why we wanted to do validity first and then go to reliability. And we're, you know, repeating that again today just like we did in December.

So are there any issues that we want to discuss? Does everybody feel like you understand what was provided by the developer, the new information and what that seems to tell us? You understand why we at NQF reversed our initial preliminary rating and why we now would rate it as moderate? You understand your options in terms of what you can do in terms of the voting?

And let me go back, I know, Sean, you came in to the call late. I know one of the things that we had talked about was the thinking behind why we, as NQF, made the change. That was included in the memo in a very brief way.

But essentially, just to repeat what we said before, even though the developer didn't have the formal sensitivity specificity statistics, they were able to tell us basically, you know, what things were in the medical records that weren't in claims and vice versa. And particularly for the new exclusions that they've added to the measure, so that's why we changed from insufficient to moderate.

Sean Bryan: OK.

Karen Johnson: I don't remember if there was anything else that was discussed before Sean came aboard. We talked about table seven in the submission and just – we familiarized ourselves with how to understand that table and what that was telling us. And as you've heard, Sean, there has been more discussion that I think started before you came about the – it seems that the trauma exclusion and the neurologic impairment exclusion from the early 2000s data, it looked like they weren't being captured by claims.

But since then even though there's not new information that specifically speaks to that, there's at least some feeling that that may not be the case. If you went – if you were able to go back and do that testing now that perhaps might not be the case given the lack of comments in their user system – you know, by users of the measure. And also, the fact that even though the number of patients increased with the new specifications, the actually number of exclusions decreased.

So that gives a little bit more light on what might be going on again with that change from the trauma being a look back period of one year. It's now a look back period of three months. I think I covered everything. If I missed something, can someone remind Sean and me of anything else that was discussed?

Sean Bryan: Can you clarify because I share Robert's concern about the neurologic impairment only 25 percent being picked up by the administrative data and it was referred to as a small number. It's 25 percent of what?

Robert Chou: Their total number is 4 percent, so they were – they were capturing – 0.9 percent of patients had neurologic impairment based on administrative data, but it was 4 percent when you look at the medical record data.

Sean Bryan: Got it. OK. Thank you.

Karen Johnson: And Sean, not to put you on the spot, but did you have any other questions or anything else that you wanted to ask that perhaps was discussed earlier? But we want to make sure that this is a fair vote that everybody understands and had access to all the information.

Sean Bryan: No. Not at this time.

Jason Matuszak: This is Jason. I just want to refresh my memory. How long are these measures once – if they are approved, how long are they approved before they have to come back to the committee for re-analysis?

Karen Johnson: On average, it's about three years. One thing that we have a potential to do – the developer, as they mentioned, they won't have updated data until the summer of this year. They potentially could bring back, say, in a year. When they do their annual update we could ask them to bring back updated reliability testing.

So again, we're really talking about validity and having to base, you know, your information on the validity data that's available. But potentially, in a year, if you wanted to, if it goes through, you could actually ask them to bring back updated testing data at that time. We're pretty sure – and (Jenna) correct me if I'm wrong – but we're pretty sure they could do updated reliability testing.

But as she mentioned earlier, she might not be able to do a different type of validity testing. The score-level validity testing which they have not done can be difficult if you don't have other measures to correlate with. So there would be no guarantee that in a year they would have anything to show you in terms of validity. And, (Jenna), is that correct?

Jason Matuszak: I guess the other – I guess the other part to my question before (Jenna) answers is then – does it then also follow that in three years' time that when they – this measure if it gets approved now, if it comes back in three years' time they will still not again have any better validity data than 2002 data?

Karen Johnson: Well, there's a couple of things. I mean, if it goes through this time, you certainly as a committee could very strongly recommend that you do want to see some additional information in terms of validity and data element validity. You could certainly do that.

And then, you know, three years down the road where you can see whether NCQA is able to do that or not, I think that's probably – all I can say about

this – I mean, you know, things always could change in terms of what we expect. Right now, we do not require developers necessarily to update their testing. But if it's a strong recommendation for the committee, that would be signaling to them that you, you know, might not be willing to let it go next time if they couldn't do something.

Kimberly Templeton: And this is Kim. I would have a concern about approving anything that again has data that's 15 years old. So it would be great if they could get us data the next time, it would be even better if we would have data now to make that decision.

Karen Johnson: Yes. I think it's pretty clear that they can't do it now. So, you know, that – it's a difficult position. I think you have to – you'll have to vote in what you – what you feel and, you know, we really think this is a valid measure. Is anybody being – you know, this is a fair measure is what we're trying to think, you know.

(Jenna): Yes. Karen, this is (Jenna). So like – yes, we will have – plan to submit their data in June, so we would be able to do reliability testing later this year. I mean, I guess, as far as thinking about other types of validity testing, we could potentially see if there's an integrated plan who'd be able to help us to look at the difference in claims and maybe what's in record for the – for the data elements.

You know, I think that would – that could potentially take some more time to find plans who'd be willing to participate in that. But that is something that I guess we could potentially do rather than having to actually look, you know, at paper medical records, which, as I mentioned before, could be really resource intensive. But potentially and with an integrative plan maybe that had EHRs we could potentially test it for future – for the future.

Karen Johnson: Thanks, (Jenna). And this is Karen. And I don't know if you guys had a chance to look and I have no idea if you looked if it would exist. But NQF also allows pulling that kind of information from the literature.

So if someone else along the way has done any kind of look at, you know, accuracy of claims data in terms of trauma compared to, you know, medical

record, you could – you know, you could bring that to bear and that – you know, it would be a different sample and different timeframe and etcetera if that exists. That would be allowable.

(Jenna): OK. Thanks, Karen. Yes. It was not something we were able to do this time, but maybe something that we could do for the future as well.

Karen Johnson: Any other questions from the committee? And, (Jenna), any other parting thoughts before we go to votes?

(Jenna): No. Thank you, Karen, for giving us the chance to walk through all of this today.

Karen Johnson: OK. Let's make sure we still have 13 voting members. Sean, nobody has left us or dropped off accidentally hopefully?

(Shawnn Bittorie): Not that I see.

Karen Johnson: OK.

Sean Bryan: I'm still here.

Karen Johnson: OK. All right.

(Shawnn Bittorie): And then, how do we want to handle the votes for Sean Bryan?

Karen Johnson: Sean, are you comfortable giving a verbal vote or do you want to try to do a chat and let our folks behind the scenes add it up for us? What do you prefer? It's totally up to you.

Sean Bryan: I'm OK with a verbal vote.

Karen Johnson: OK. Thanks. Sean, do you want to walk us through the voting process? It sounds like we're ready to go ahead and vote.

(Shawnn Bittorie): Absolutely. So in just a moment the NQF team will advance the slide to the slide that will have the voting options. You'll see small boxes next to each of the options.

Voting members only will click in the boxes next to the answer of their choice and voting option of their choice and the results will populate in real time. If the boxes don't show up for you, you can refresh your session by pressing F5 on your keyboard for PC and or command-R for a Mac.

Karen Johnson: Thank you, Sean. And just one thing, we made a little mistake with this slide. Disregard high as an option. Your options will be moderate with a two, low with three or four insufficient.

(Katy Streeter): Thanks, Karen. So now voting should be open for 0052. We are voting on validity. Two is moderate, three is low and four is insufficient. And Sean?

Sean Bryan: My vote is low.

(Katy Streeter): OK. So we have five for moderate, four low and four insufficient.

Karen Johnson: And post comment.

(Katy Streeter): OK. And because there is no gray zone during the post-comment evaluation, this measure will not pass the validity sub-criteria.

Karen Johnson: So this is Karen. What this means is the measure has not been recommended by the committee because it did not pass the must-pass criterion of validity. At this point, we won't bother with the vote on reliability and that's it.

Committee, is there anything you would like to suggest for NCQA for future? Would you like to see them bring this back when they do have additional data? Is there something you'd like to see?

Roger Chou: I mean, this is – this is Roger. And, you know, I think, you know, we all – I think everybody thinks that this is a worthwhile issue and that we want to reduce overuse of imaging. I think having updated data on the validity, particularly the exclusions and understanding – I mean, I would be much more comforted if we had data showing that, you know, the rate of traumas that are missed is, you know, 5 percent or whatever, not 20 percent. And – you know, and not reliant on 2002 data I think is the other issue.

So, really, I think to have at least something – I mean, it's really hard, for me at least, and maybe the other committee members to approve – to vote something as moderate or high validity when the actual data says that it isn't. You know, regardless of what the plans are saying, I mean, the actual data says it isn't valid.

So I would really like to see, you know, more updated data. And again, I think everybody is fully in support of the general concept. We just need to make sure that the measure is, you know, doing what it's meant to do.

Karen Johnson: Thanks. Anybody else from the committee, any thoughts or does that cover it?

Sean Bryan: This is Sean Bryan. I concur with Roger.

Karen Johnson: Great. Well, thank you folks from NCQA for being on the call. (Katy), let's talk about the next measure.

(Katy Streeter): Thanks, Karen. Moving on to measure 0514, lumbar spine for low back pain. The measure is a CMS measure. This measure did not pass the validity sub-criterion last time. It was not recommended for endorsement. We did receive five comments on this measure, two comments – two comments supported the measure and three comments are – supported the committee's recommendation not to endorse the measure.

We also did receive a request for reconsideration from the developer and, actually, before we get into the standing community discussion about this, we'll be asking you to determine whether or not you would like to reconsider the measure. And then, we'll walk through – if so, we'll walk through the remaining criteria and have you vote again. But before that, we'd like to pause and give CMS the opportunity to speak to their reconsideration request.

Colleen McKiernan: So this is Colleen McKiernan from The Lewin Group. Can you hear me?

(Katy Streeter): We can. Thank you.

Colleen McKiernan: Great. OK. I'm going to just give the – a quick introduction to NQF number 0514 on behalf of CMS, Yale CORE and Lewin. So thanks again for the opportunity to speak with you all about NQF number 0514, MRI lumbar spine for low back pain. On behalf of CMS, Yale CORE and its partner, The Lewin Group, work to maintain NQF number 0514 a measure originally endorsed by NQF in 2008 and in use within the hospital outpatient quality reporting program since 2010.

As a reminder, the initial patient population for NQF number 0514 includes lumbar spine magnetic resonance scans with the diagnosis of low back pain on the imaging claim. We then removed MRI studies from the denominator for patients with a history of a red flag condition.

Examples of these red flags can include things like cancer, HIV, trauma and other diagnoses. Of the imaging studies included in the denominator, those for which anteceded conservative therapy was not performed, so that's patients who underwent an MRI lumbar spine study for low back pain who didn't have claims-based evidence of chiropractor, physical therapy or evaluation in management proceeding the imaging study are captured in the numerator.

This measure is designed to promote conservative management prior to MR Imaging in lower risk patients specifically those with uncomplicated low back pain. And the measure is specified with substantial technical and clinical input and, today, it's one of the few measures of imaging efficiency that's endorsed by NQF and used in public reporting without report of unintended consequences or harm. We haven't heard anything from stakeholders about either of those.

On March 3rd, CMS, Yale CORE and Lewin submitted a request for reconsideration for NQF number 0514. Based on NQF's measure evaluation criterion guidance, we believe that this measure aligns with the moderate validity recommendation in the NQF algorithm guidance for steering committees as it has received in prior reviews for endorsements.

The measure's specifications are aligned with the most updated clinical practice guidelines and have strong face validity. Additionally, measure testing confirms that threats to validity have been addressed by the exclusion of red flag conditions.

NQF number 0514 also did pass the importance and reliability criteria during the most recent review back in January. As one standing committee member stated during the review webinar, there will always be exceptions in health care. And as long as the rate of exception is low, performance scores will not be impacted and the measure serves its purpose. We believe that as currently specified, the measure identifies clinically significant variation and is an important tool for CMS to promote hospital-quality improvement for use of (inaudible) services.

So I know that some members of the standing committee have some suggestions on diagnosis that should be added to or removed from the said specifications. We welcome this feedback and we'll bring it to our multidisciplinary TEP when we meet with them later this year for the measures annual review.

We do seek to improve the measure every year. CMS is committed to continual improvement regardless of whether the request for reconsideration is heard today or not. So we look forward to today's discussion and appreciate the opportunity to speak with you all once again about this measure. Back to you.

(Katy Streeter): Thanks, Colleen. Comments from the committee?

Kimberly Templeton: This is Kim. So I guess my question is has there been – have there been any substantive changes to this since the committee last voted against this measure?

Colleen McKiernan: So we took the feedback that you all provided and that was summarized in the memo and went through in kind of detailed in the request for reconsideration. But we responded to each of the comments.

In order for us to make changes to the measure specifications, it's a longer process. And so, certainly if there are – if there are items for which you have a feedback that you say this needs to be changed before we could approve it, we would need to bring it to our TEP. But we have detailed – we have responded to each of the major things that were mentioned in the memo and in the transcript in our request for reconsideration to kind of clarify exactly where we stand on things since I know we went back and forth on a lot of issues and we kind of went in circles during the last call.

Kimberly Templeton: OK. So – but it's the same measure that we declined to endorse the last time, so we're just re-voting on what we've already decided. That's just a comment. Not a question.

Colleen McKiernan: OK.

Karen Johnson: And this is Karen. Just a clarification. The developer has asked you to reconsider, but you don't have to reconsider. So you actually don't have to re-vote.

So part of the discussion should be, you know, do you feel like that the developer in their letter or in their request for reconsideration, did they – did they convince you that maybe you would change your vote if you were allowed to vote again. And if so, then I think you should ask for a re-vote, but we're not – you don't – you're not required to re-vote.

John Ventura: This is John. I went back and looked at the NQF's criteria for the committee to look at judging validity. And if you follow the algorithm, it does meet the standards for face validity and that recognized experts agree that it is able to separate good from poor quality care and they showed the data. I think it was 82 percent agreement or strongly agree that it would do that.

And then, the next part of the algorithm is that there have to be no potential threats. And again, I think they were able to show that adding additional exclusions wouldn't appreciably change the scoring. And, therefore, it meets the moderate standard of evidence for validity.

Roger Chou: This is Roger. Maybe we could – I think Craig, Thiru and Katherine were supposed to be the lead discussants and maybe I'd like to see if they want to summarize kind of their thoughts?

Thiru Annaswamy: This is Thiru. I can jump in. So some of the comments that were made during the January call related to how the cited guideline was used in specifying this measure, some of the comments in the – on the call related to how diagnosis codes used did not necessarily reflect the uncomplicated low back pain condition that was cited in the – in the guideline.

So, as such, the validity of the measure which is based on looking at uncomplicated back pain instances and, in those instances, conservative care – anteceded conservative care would be very important and appropriately backed by literature before ordering MRI. The claim that we're trying to limit trigger happy clinicians from ordering MRI at the first visit without giving adequate trial of conservative care may apply to uncomplicated back pain cases. But the way the measure – the diagnostic codes and criteria are used is not really reflective of the literature side.

That was in my mind was the primary reason why this measure did not pass validity at the January call. So, I'm looking back at the quoted ACR, appropriateness criteria, and I'm going through the different variants one by one. The first variant is acute low back pain, sub-acute low back pain, chronic low back pain with or without radiculopathy and no red flags and in that it's usually not appropriate.

The second variant is low velocity trauma, osteoporosis, elderly individual, chronic steroid use and those are where it's usually appropriate. But then, when you come to – I'm sorry – and then, variant three also appropriate which is immunosuppression, cancer, infection. Some of the areas where it's considered appropriate is variant four where you have surgery or intervention that is being considered or the patient is a candidate and, in such cases, an MRI would be usually appropriate.

And the – in this variant, the description was acute, sub-acute chronic low back pain or radiculopathy within six weeks or during six weeks of

conservative care. And the criteria says it may be appropriate. And then, variant six is the rapidly neurological deficit and variant five is prior history of lumbar surgery with new or progressive symptoms. So that's part of it and the other part of it is what I mentioned earlier about the diagnostic codes. So those are sort of the two main points I felt and I'll stop there.

Roger Chou: Katherine or I think Craig wasn't on the line, but, Katherine, do you have additional comments?

Katherine Gray: Well, I have more of a question. The answer to Kim was that there's nothing different about this. I interpreted the (LOU) and letter to be an argument for the re-vote. Would somebody from the developer team make some of the key points that's in that, or do they?

Colleen McKiernan: So this is Colleen – sure, this is Colleen McKiernan. I can give a high-level summary. There's a lot of content in there, but I'll try to hit on the key points.

So in the letter that we submitted, we hit the three sub-criteria for the validity criterion, so measure specs, validity testing and threats to validity and we walked through why we think we met each of those sub-criteria including some of the comments that were made about different codes as was previously mentioned for inclusion or exclusion both in the definition of uncomplicated low back pain as well as in the – in the list of exclusions.

We also go into detail in our – into our face validity assessment. That's something that we didn't have time to cover during the January call and so we want to make sure we presented while we have strong face validity. I mean, we detailed the results of that face validity assessment here.

And, finally, we walked through the threats to validity which, during the discussion, centered around a few exclusions. Things like the lumbar spine surgery and other – like look backs for certain exclusions. So, collectively, we've kind of hit on each of those three sub-criteria and why we think we've met each of them.

And, in the conclusion, we summarized overall that we think we've – that the measure is currently specified, addresses the broader patterns of care, excuse me, for appropriate versus inappropriate low back imaging. I can go into detail on specific items if you'd like, but there's a lot in the letter, so I'm happy to address specific questions as well.

Kimberly Templeton: This is Kim. I guess I still have the same concerns that we had – that I had the last couple of times we've discussed this measure.

One is in the numerator. It claims for physical therapy and chiropractic evaluation and manipulation. Those aren't the only forms of conservative therapy that one can do for low back pain. Granted those are the ones that you're going to pick up in the EMR or in the clinic notes, but that's not the only thing that can be used. So I think we're – other people – people are going to be using other conservative modalities that we may not be able to pick up which I think then is a real risk to this measure.

My other concern that, on exclusions, I think the look-back period is too short especially those with – who've had prior surgery and those who've had cancer. People having back issues with those – with those histories don't have it within a brief period of time even within the succeeding 12 months. It may be years down the road. I think those – and those, as I mentioned before, do not constitute uncomplicated low back pain.

(Dr. Bruetman): Can I clarify – this is (Dr. Bruetman) from the Lewin Group. Well, I understand exclusion as a period, isn't this – we're trying to develop measures with the available feasible data and one could – while I agree that we cannot find an exclusion going back that much time, wouldn't that be then therefore no valid measure because we would not have a patient in their 50, 60, 40, whatever age you want, and they have cancer when they were young – there're two options. Either we'll never find it or it would be probably – if there are some relation to it, it should be in the claim because the doctor appropriately would say we are asking MRI for a patient for his low back prior cancer X, whatever cancer.

So if now we are assuming that anybody could have had that and we will never find that exclusion and that's going to happen with every single measure. And I think one of the goals is to have measure that are feasible and reliable, but also that we can eventually have some measures for doctors to provide appropriate care.

And I think we are – with those limitations we would have any case that would never work for any measure because we wouldn't have it unless we accept that if we go and look back for five years, if it hasn't been in the last five years any claims of that, we should assume that that's not the case. And we do a look-back period, and Colleen you can clarify, for cancer it's five years, correct?

Colleen McKiernan: So for cancer it's 12 months that was set back in – that was set back when the exclusion was originally added in 2012 for both the lumbar spine surgery and the cancer look back. These were decisions that were made by some recommendations from our expert panel.

Obviously, if there's – it's like the sword on which we're going to fall is the look back for an exclusion, that's something that we could bring back to our expert panel and discuss whether a longer look back would be appropriate.

Obviously, having a longer look back gives kind of the benefit of the doubt to providers looking back, therefore, that cancer five years ago was a – there was a suspicion of metastasis, for example, in the lower spine region. And so, that's something that we can definitely bring back to our expert panel to talk about extending the look backs as far as claims are available.

(Dr. Bruetman): So wouldn't it be appropriate to have it on the claim if they're looking – if the doctor is looking specifically for that back pain for a person that had cancer, it is (probably) that we would either have that on the claim or, you know, not have it. Furthermore, we also want to clarify we are not expecting the measure to be zero. If there would be zero, we always had to have expectation, but we understand that.

What we – I think the intent of the measure at its – at the end of the day is, you know and it's – you know, we always use the same term, maybe not the

most appropriate, but it's the – what we call the fast trigger doctor that patient comes in, low back pain and immediately gets, OK, go and get an MRI. That's the goal.

So I think by doing some look backs and doing, you know, the 30-day or, you know, how we do the claims and it's done immediately that's what we're trying to avoid on the patient. And then, we add all these exclusions that really clarify all the issues that why a doctor would make a decision to immediately or in the next week or whatever to the MRI or request MRI, you know, they had HIV, cancer or some recent surgery et cetera.

Roger Chou: So this is Roger. And, first, just to address the – you know, whether cancer is coded as a reason for doing an MRI and the answer is no. There is no code for that. When I – when I order an MRI, it's for low back pain.

You don't code history of cancer typically as part of the diagnostic code. That may be put in the – you know, in the reason for getting the MRI in kind of text or whatever, but that's not – I mean, I would say that the majority of the time it is not coded. And the same – and I think it's even more of an issue when we're concerned about cancer for other reasons like somebody has reports they've had some weight loss or whatever.

All those constitutional symptoms that we talk about, those are not captured first of all in the denominator exclusions. And I think that they're unlikely to be coded as well. If we had – if you had data to show us that that stuff is always coded or coded most of the time, that would be great. But I haven't seen that data.

And the other issue with the cancer thing in particular is that late metastasis really are a concern particularly with things like breast cancer which can occur – you can get late recurrences of breast cancer, are actually quite – you know, it's not uncommon, unfortunately. And cancer is one of the – a history of cancer is one of the major red flags.

So, I mean, you know – you know, there – I totally agree with what's been said about, you know, if there are small proportions or whatever, then that's reasonable, but we haven't seen data to show that. And again, there are a lot

of – you know, many of the times when we're doing the MRI it's because we're worried about a condition, not because the condition is, you know, necessarily there.

When somebody has a history of prior cancer, their probability of having cancer on, you know, MRI is 10 percent. Ninety percent of the people will never have cancer, you know, when we – when we do the MRI and that's OK. I mean, we have to do a lot of the, you know, testing to capture those few patients that are going to have, you know, a serious condition. That's what we actually expect.

But again, some of the – some of the way that these specifications have been laid out I think has – have been concerning because they don't really address it, you know, the way that the guideline say should it be addressed, number one; and also the way that it's done clinically and how things are coded and practiced.

And I'll just echo what Kim said about the – you know, if you look at – if you look at the most recent ATP guidelines, there's a lot of things that they recommend as, you know, first line treatments for low back pain including psychological therapy and all sorts of other things. And none of those are captured in the conservative – you know, in the types of conservative care that you want people to have. So I think that that's the other major issue.

Kimberly Templeton: And, Roger, I would echo everything that you've said regarding cancer. I'm also an – I'm an orthopedic oncologist. The most common time for recurrence in bone is when it's in the first two years of diagnosis, but those can occur up – it can occur up to 10 years. And so, even looking at the 24-month period, that's not going to be picked up in the current specifications let alone those people that recurred 5 years and 10 years out.

Colleen McKiernan: So this is Colleen McKiernan just to respond to that briefly. So we do look for all of our exclusions when we have the look back period. We're looking at every encounter that a patient has over the look back period time.

So if a patient has had – a diagnosis had a – was treated for cancer two years or three years out, if they go in for follow-up care just to a yearly scan or

however their oncologist is treating them, we would capture that follow-up care. So even if the diagnosis was several years out, so long as they're having one encounter at which there's documentation in the claims that the patient had cancer or HIV or any of the other exclusions, then we would capture it. So, we're – we are looking everywhere in the inpatient, outpatient and physician data to try to identify the exclusions.

Christian Dodge: This is Christian Dodge and I just wanted to clarify also that there are codes that indicate a personal history of cancer that probably should be in the record. So these codes do exist.

And I think Roger's point about whether they actually do get coded is a different question, more based on provider performance or, you know, kind of how thorough they are in representing the true nature of the case. But the tool to do that even if you aren't actively treating the cancer to be able to provide that kind of a code is in the code set.

Roger Chou: Yes. Thanks, Christian. That's correct. I was referring more to kind of common practice which I don't think is necessarily to code all that stuff in.

And, Colleen, that was – that was helpful to understand that you guys are looking at all the codes, not just the ones connected to the imaging visit. But I – you know, but again, it would – for me, it would be much more compelling to be able to say that if somebody has a history of cancer to show data saying that that gets captured if you do a one-year look back or whatever the time period is.

I haven't seen that, so, you know, I don't know how many people are not – you know, I mean, there might be people who have – who have breast cancer, you know, 10 years ago or six years ago or whatever it is who, you know, I don't know what the proportions of those are that are going to have a code, you know, at year six or year five. So, you know, if we – if we had that information that would be very helpful for me.

(Dr. Bruetman): Can I clarify? So I don't disagree and we're going to miss – there are two concepts of that and this has been a challenge with measures in general with

claims is there is any show of – you know, not having information on the claim and there's people that put more, people that put less.

So the question is if the hospital in this case is not providing all the information, would that – shouldn't that be an issue, like, well, you know, for people to be able to measure performance well that becomes a problem and a burden on the hospital to do it right. And not just to do it right and go off – to up code but to do the right thing.

And if they don't do the right thing, well, they will be – if they didn't have the right information in there because they didn't put the history of cancer that they could've, then there should be – like that shouldn't get them off, well, then you're OK. It should be like, well, learn to code and do the right thing. And one could say well, since they're not paid for this, you know, they wouldn't – they don't have to code it. But, you know, they wouldn't have – that's another issue to consider.

The other issue is, again, I just want to insist on this concept that we – I know that we talked – we explained that we look at all the patient history not just the related to the claim and that visit, but all the patient's history. So we look back at the patient and if they had another – an oncology visit and they had, you know, any other things related.

So there could be cancers, it's true. That may not appear for many years and we don't have to look back, but that's going to happen with any case or you could even say for a cancer, the one they were younger. We will never have that for any measure that is related to cancer that has a cancer look back period requirement or a red flag. We'll never have that.

The second is there are going to be missed cases because, yes, they might not have been in the last three years or two years et cetera, so they're less frequent, but you might miss those. And we understand that and that's why, again, I want to say this, it is not zero the expectations. But I don't think we can accept the measure that has a 30-something percent and it's because we believe that, well, probably 80 percent of those are maybe those cancers that we missed.

Probably there's something – I think we all agree that there is problems to the literature that there is an overuse of MRI for low back pain. It's in choosing wisely for a reason. It's been looked by medical societies. You know, it's been in constant discussions on providers of the overuse.

So while we may miss cases that are – maybe should have been excluded for the numerator, that should not make the – to get the perfect measure should not stop us from having a good measure that helps providers do the right thing. And I think by not having a measure today in the – in NQF or that is endorsed that says MRI lumbar spine should not be a common use, we are missing on an opportunity to improve care with all the (potential) consequences.

Kimberly Templeton: But some of that including care includes using – utilizing all conservative modalities. We're really not encouraging providers to do that if the only thing that we have listed are chiropractic care and physical therapy. I think the real goal, you're right, is to decrease use of MRI, increase the use of conservative modalities. But this measure is not going to inspire people to do that because we don't have all recommended conservative modalities included.

(Dr. Bruetman): We have bought all E&M visits, so it's – I mean, there could be – it's true, but visits that are done to the doctor's office if we're doing the timeframe are included. So we have a number of issues that we might not be able to get certain modalities, but it is pretty covered. I mean, a lot of the things are covered and some are never going to be covered, that's true. There are going to be issues that are never going to be covered. We recently can't code them.

Colleen McKiernan: Yes. This is Colleen McKiernan again. I know that the American College of Physicians recently, in February, released recommendations for conservative therapy and treatment of low back pain. And this is certainly something that we'll be bringing back to our expert worker, or I'm sorry, our TEP.

As you know, there's not an immediate change in specifications when new literatures are released. And so, it's definitely something that I personally have flags to discuss with our experts once we do meet with them later this

year. But it is on our radar that there might be updates that are flagged here in this guideline as well as in other recently published evidence that we want to make sure we capture if it's appropriate based on our experts' feedback.

(Katy Streeter): Thanks, Colleen. This is (Katy). I think that what we would like to do is ask the committee if there are any members who do feel that the committee should revisit and re-vote on this measure beginning with the validity sub criteria.

(John Ventura): Are you – I'm sorry. You're asking us to rethink it deserves a revote?

(Katy Streeter): Yes. That's correct.

(John Ventura): Then I would say yes based on my prior comment that it meets the standard for face validity.

(Katy Streeter): OK. So what we'll do then is we will have the committee re-vote on this measure. Like I said, we will start the sub-criterion where the measure failed last time which is validity. Also Sean, can we do a quick check to make sure that we do have our quorum.

Sean Bryan: Yes. I'm here. But just one member of the committee said that they do think it's appropriate for a re-vote. What are the rules on the committee voting to re-vote or not?

(John Ventura): I didn't think I had that much clout, Sean, so.

Kimberly Templeton: I thought it was more than two, not just one.

Karen Johnson: No. This is Karen from NQF, we don't vote on whether we need to re-vote or not. Basically, if there is – you know, even somebody on the committee that, you know, feels like it's worthy of another look just to go all out, we will have the vote and we'll see where it goes. That seems the fair way to do it to us.

So – but before we do that and while our com partners are checking to make sure that we still have our 13 that we require, is there any other pieces related to validity that you feel like haven't been aired today or on the last call? So not to – not to revisit things that have already been discussed several times, but that would help you in your vote today? OK.

Jason Matuszak: This is Jason. I have a question about the – you had – you had said with the last vote that we just did not because it was a – based on the comments that they were doing the voting that there was no gray area, so does this kind of reset where we were with this measure previously? So does it go back where we were before or is this considered post comment?

Karen Johnson: So this is post comment. So just like the other one, this is – this is your final say as a committee unless something happens later in the process with our (CSAT). But basically, today, what we'd be looking for is a vote on validity and you would have to actually vote – more than 60 percent of the folks on the call would need to vote moderate or high – actually moderate is your only option, (Katy), correct me if I'm wrong on that one.

But we need more than 60 percent of you to vote either moderate or high or the measure would fail and would not be recommended by the committee. Did that answer your question? I'm not quite sure I.

Jason Matuszak: Yes. Sounds – that's good. Thanks. I just didn't know if we reverse time or just moved ahead, yes.

Karen Johnson: Right.

Jason Matuszak: That's good.

Karen Johnson: OK. Anybody has any other issues that you feel like haven't been adequately addressed? And we're hurrying along just a little bit because we're looking at the clock, so apologies for that.

(Shawnn Bittorie): And Karen, as you know, Craig stepped away for a minute. If he's back, we're still at 13, everybody else is – we still have 12 logged in and we have Sean on the phone.

Karen Johnson: OK. Craig, are you there?

(Shawnn Bittorie): His line is still dialed in, but I think that's a no.

Karen Johnson: All right. Well, this is – this is unfortunate that we've lost the quorum, so we can't actually have the vote. Going once, going twice, Craig, if you're there, flag us in some way. OK. We don't have quorum so we unfortunately cannot do the re-vote. We will do as we said we would do. We'll send out a …

Kimberly Templeton: This is Kim. Is there any way we could just do a preliminary vote to see if one vote's going to make a difference one way or the other or if there's already an overwhelming majority going with one way or another? At least we'll have some indication.

Karen Johnson: You mean just to vote unofficially here as our 12 and just see where we're landing? Is that what you're asking?

Kimberly Templeton: Right. Because if it's overwhelmingly one way or another, we could have Craig vote, but it's not going to change the outcome and at least we'll have some idea of where this is heading.

Karen Johnson: Yes. I'm trying to think of the pros and cons of that. Let's go ahead and do that. So we'll have a non-binding vote just to get the flavor of where the committee is now and, (Katy), are we ready?

(Katy Streeter): You should be able to vote now, there 0514 validity. And, Sean, if you would like to verbally give us your vote, we'll add it in here as well.

Sean Bryan: My vote is low.

(Katy Streeter): OK. So it looks like we do have 12 votes including Sean's. Five voted moderate, five low and two insufficient.

Karen Johnson: And again, this is non-binding. It looks like if this (really) vote right now, it would not go through. And my brain is not going to be able to do the math right now to tell us now if Craig were actually on the line if his moderate would tip us over.

Kimberly Templeton: No. It would – it would – it'll still put us at seven low to insufficient and six with moderate if Craig voted moderate. So it still wouldn't go through.

Karen Johnson: It's still under 60? OK.

Kimberly Templeton: Yes.

Karen Johnson: All right. Right now, I believe this would still be – this is still an unofficial vote. I will check with our more senior staff to see if they want to overrule me and make this an official vote even though we technically don't have quorum. But I suspect that we will be sending the SurveyMonkey out to be official with this.

The one caveat that I will say is if – when we – if we do send out the SurveyMonkey which again I think we will, we will have to ask you on the SurveyMonkey to vote for Sean's validity again just like you did. We'll also have a feasibility and usability vote, as well as an overall up or down vote.

However, if the measure actually does officially go down on validity, we will vacate the up or down vote. So, in other words, we just won't even look at that because if we were able to do it live, if it goes down on validity, we don't go forward. So apologies for that kind of technical work around that we have to do there.

With that, looking at the clock, it's getting very close to time. I think you've provided some feedback – some good feedback to the developers. Again, this is still a little bit preliminary, but I'll ask you very quickly is there any kind of advice you'd like to give to the developers if this does in fact officially go down, any advice from you guys that might be in addition to the things that you've already mentioned today? And – OK.

With that, I'm going to hand it over to (Katy) to finalize the call. Thank you guys so much for joining us today.

(Katy Streeter): Thanks, Karen. We do need to open up the lines for public comments to give those out there the opportunity to make a comment if they would like to. Operator, if you could – go ahead.

Operator: At this time, if you would like to make a comment, please press star, then the number one on your telephone keypad. We'll pause for just a moment. And there are no public comments at this time.

(Katy Streeter): Thank you. So as far as next steps …

Operator: Excuse me.

(Katy Streeter): Yes?

Operator: You do have a public comment from (Heidi Bossley).

(Heidi Bossley): Hi. Can you hear me?

(Katy Streeter): We can, (Heidi).

(Heidi Bossley): OK. Great. Sorry, I wasn't very fast with the star one. This is (Heidi Bossley), speaking on behalf of the Federation of American Hospitals. I just wanted to say that we appreciate the thoughtful discussion that all of you have today and support the committee's concerns over validity of both measures.

We think it's very important that not only should we be looking at the time that it takes for developers perhaps to test and maintain this. We recognize that, but it is also expensive to implement measures and endorsement should not be continued on measures that are not scientifically bound nor if they do not produce results that we believe improve clinical outcomes or fairly compare providers, they should move forward. So we are very appreciative of the discussion today and support the final decision of the committee. Thank you.

Karen Johnson: Thank you. Any additional comments?

Operator: There are no further comments.

(Katy Streeter): Great. Thank you. So, for next steps, like we said, we will most likely be e-mailing you the committee a survey link to do the official voting. We will be in touch with both the committee and the developers regarding a firmer timeline of the CSAC review dates and so on.

So any other – any other comments from committee members before we end the call today? OK. Well on behalf of the staff here at NQF, thank you all for joining us and we will be in touch soon. Thank you.

Kimberly Templeton: Thank you.

Male: Thank you. Bye.

Male: Thank you. Bye.

Male: Thank you.

END