NATIONAL QUALITY FORUM

+ + + + +

MUSCULOSKELETAL MEASURES STANDING COMMITTEE

+ + + + +

WEDNESDAY
MAY 7, 2014

+ + + + +

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Roger Chou and Kim Templeton, Co-Chairs, presiding.

PRESENT:
ROGER CHOU, MD, FACP, Co-Chair
KIM TEMPLETON, MD, Co-Chair *
THIRU ANNASWAMY, MD, Dallas VA Medical Center
CARLOS A. BAGLEY, MD, FAANS, Duke University School of Medicine
STEVEN BROTMAN, MD, JD, AdvaMed
SEAN BRYAN, MD, Greenville Health System
CRAIG BUTLER, MD, MBA, CPE, American Academy of Orthopaedic Surgeons
KELLY CLAYTON, BS, Arthritis Foundation
LINDA DAVIS, BSN, Minnesota Health Action Group
JAMES DANIELS, MD, MPH, FAAFP, FACOEM, FACPM, Southern Illinois University
CHRISTIAN DODGE, ND, Bastyr University
ZOHER GHOGAWALA, MD, FACS, Tufts University School of Medicine
V. KATHERINE GRAY, PhD, SAGE Health Management Solutions, Inc.
MARCIE HARRIS HAYES, PT, DPT, MSCI, OCS, Washington University in St. Louis School of Medicine
MARK JARRETT, MD, MBA, North Shore - LIJ

Health System
PUJA KHANNA, MD, MPH, University of Michigan
WENDY MARKINOVICH, BSN, MPH, RN, Blue Cross and Blue Shield Association
JASON MATUSZAK, MD, FAAFP, CAQSM, RMSK, Excelsior Orthopaedics
CATHERINE ROBERTS, MD, American College of Radiology
ARTHUR SCHUNA, MS, RPh, BCACP, American Society of Health-System Pharmacists
JOHN VENTURA, DC, American Chiropractic Association
CHRISTOPHER VISCO, MD, Columbia University College of Physicians

NQF STAFF:
HELEN BURSTIN, MD, MPH, Senior Vice President for Performance Measures
ANGELA FRANKLIN, JD, Senior Director, Performance Measurement
KAREN PACE, PhD, MSN, Senior Director, Performance Measurement
ANN PHILLIPS, Project Analyst
KATHRYN STREETER, CHES, Project Manager

ALSO PRESENT:
JOHN FITZGERALD, MD, American College of Rheumatology
MELISSA FRANCISCO, American College of Rheumatology
RACHEL MYSLINSKI, American College of Rheumatology
JINOOS YAZDANY, MD, MPH, American College of Rheumatology

* Present by teleconference

TABLE OF CONTENTS

P-R-O-C-E-E-D-I-N-G-S

8:35 a.m.

MS. STREETER: Good morning. I'm Katie Streeter, Project Manager for this project.

I'd like to introduce Ann Phillips. She's our Project Analyst.

Our Senior Director, Angela Franklin, is on her way.

So I'll start us off here today.

The restrooms are out past the elevators on the right. We will be breaking today. These are actually incorrect times. We had to switch our agenda around a bit. We will be taking a break -- oh no, they are correct -- at 11:00, lunch at 12:45 and then another break at 3:30.

Helen, do you want to do -- okay.

MS. BURSTIN: Good morning everybody, I'm Helen Burstin, Senior Vice President here at NQF. Ann Hammersmith is our General Counsel and I suspect she's in

traffic, so we'll go ahead and I don't want to delay you guys.

So what we do is we actually combine introductions and disclosures of interest just for efficiency. And the idea would be that we would ask you as part of the introductions, to introduce who you are and where you're from and then include for us any information about your disclosures, particularly your grants, your contracts, anything that might be relevant to the measures before the committee today.

We don't need to hear your full CV. That would take the morning. I've read them all. We've seen them, so thank you. That's why you're seated here, so just -- brevity is fine but the major issue is to make sure that as the measures come up, people feel comfortable that everybody's disclosures are on the table and you've all had a chance to ask each other any questions of that.

So perhaps we'll begin with Roger.

CHAIR CHOU: Thanks.

I'm Roger Chou. My background is internal medicine. I'm at the Oregon Health and Science University. I direct the Evidence-based Practice Center over there.

In terms of my conflicts and most of my funding, or almost all of it, is from the Agency for Healthcare Research and Quality. I also have gotten some funding from the American Pain Society, the Yale Open Data Access Project, a few other things, but none from industry.

I have also worked on guidelines, so I've been on the ACP Clinical Guidelines Committee. I've worked on guidelines for the American Pain Society and we do systematic reviews for the U.S. Preventive Services Task Force and I think I'm here representing ACP.

Thanks.

MS. BURSTIN: Actually before you move on, just one more quick thing since Roger mentioned he's here representing ACP. He's

actually not. He's here representing Roger. He was picked for his expertise.

So when you sit at this table -- you've been nominated by others. But when you sit at this table, you're here as an individual expert. You don't have to represent the views of who you come from. We really are relying on your personal expertise and that's why you were selected.

So sorry to interrupt.

Craig?

DR. BUTLER: Okay, thank you. Hi, I'm Craig Butler. I'm here nominated by the American Academy of Orthopaedic Surgeons where for the past few years, I've kind of tried to lead our efforts in this area and we're kind of relatively nascent in the area of performance measurement development.

I co-chair the Orthopaedic Quality Institute and I chair the Healthcare Systems Committee.

By way of background, I'm a sports

medicine fellowship trained surgeon but I've kind of walked away from clinical practice in the last few years to do more on the administrative side and most recently, I've kind of ran a systems multispecialty group, but I stepped down from that a few months ago figuring that wasn't quite my cup of tea. So I do some consulting now.

I don't have any real relevant conflicts. I'm not working on any measures currently. In fact, part of what I need to do is to learn how I can help the Academy without conflicting myself here and hope to have that discussion as we continue the rest of the day.

DR. GHOGAWALA: My name is Zo Ghogawala from Boston, Tufts and Lahey Clinic. I'm a neurosurgeon and I have no relevant conflicts.

I have funding for comparative effectiveness research from the NIH and from PCORI but no commercial conflicts of interest.

DR. JARRETT: Good morning, my

name is Mark Jarrett. I'm the Chief Quality Officer of the North Shore - LIJ Heath System which is in the metropolitan area of New York. I'm a rheumatologist by trade. I have no conflicts in terms of any funding or other measurements.

DR. HAYES: Marcie Harris Hayes from Washington University in St. Louis. I'm a physical therapist clinical investigator. I do have funding through NIH but no conflicts with the measures we're discussing today.

DR. ROBERTS: Good morning, I'm Cat Roberts. I'm a musculoskeletal radiologist at Mayo Clinic. I have no disclosures. I was nominated by the American College of Radiology.

DR. ANNASWAMY: Thiru Annaswamy, Dallas VA Medical Center, nominated by the American Academy of PM&R. I'm a PM&R physician. I also work on the Evidence Committee. I'm chairing that committee and the evidence-based practice committee there.

No conflict of interest to disclose.

MS. DAVIS: I'm Linda Davis and I'm from Minneapolis. I am a healthcare consultant and I was nominated, I believe, by the Alliance, which is the business coalition in Madison, Wisconsin.

Most of my consulting these days is with and for employers and the employer coalition based in Minnesota called the Minnesota Health Action Group. I have done some consulting for Minnesota Community Measurement which is a local measurement organization as well.

I have no conflicts and I have not received any funding.

DR. GRAY: Hello, I'm Katherine Gray and I'm also from Minnesota, from Minneapolis. I am the founder and president of SAGE Health Management Solutions and that provides clinical decision support for imaging.

And my background is Ph.D. in

Individual Differences, so testing which is reliability and validity and how you measure things. And then I did a post-doc in gerontology and the particular panel that I was on is from CMS so it all does kind of hang together.

And I have no conflicts of interest, either.

Thank you.

DR. MATUSZAK: Hi, my name is Jason Matuszak. I'm a primary care sports medicine physician, family medicine trained up in Buffalo, New York, nominated by the American Academy of Family Physicians.

DR. BRYAN: Good morning, Sean Bryan. I'm a family physician and primary care sports medicine physician. I'm in Greenville, South Carolina. I'm the chair of the Department of Family Medicine for Greenville Health System and also an associate professor at the University of South Carolina School of Medicine in Greenville.

I was nominated by the American Medical Society for Sports Medicine and I also serve on their Healthcare Transformation and Quality Committee and I have no conflicts to disclose.

MS. CLAYTON: Hi, my name is Kelly Clayton and I was recommended or, yes, recommended by the American College of Rheumatology. I have done some advocacy work with them for probably the last five years when I sat on the National Public Policy Committee for the Arthritis Foundation.

I'll finish my MPH in about three weeks, but I'm kind of here to bring a patient perspective to things.

MR. SCHUNA: My name is Art Schuna. I'm a pharmacist at William S. Middleton VA in Madison. I've spent my entire career in rheumatology practice there and I'm also a clinical professor at University of Wisconsin School of Pharmacy and I represent the American Society of Health System

Pharmacists and I have nothing to disclose.

DR. DODGE: Morning, Christian Dodge, Bastyr University, Seattle, Washington. I was nominated by the American Association of Naturopathic Physicians.

I'm a clinical professor of physical medicine in Seattle at Bastyr and also in private practice. No disclosures.

MS. MARKINOVICH: I'm Wendy Markinovich and I was nominated by Blue Cross and Blue Shield Association, the national trade association for all the Blue Cross and Blue Shield plans nationwide and I am the lead for our Blue Distinction Centers for Spine Surgery and Knee and Hip Replacement and I have no disclosures for the measures we're evaluating today.

DR. VENTURA: I'm John Ventura. I was nominated by the American Chiropractic Association. I've served on the PQRS TEP for Developing Chiropractic Measures and was part of the NCQA Back Pain Recognition Program from

the pilot project and I have no disclosures.

DR. KHANNA: I'm Puja Khanna a rheumatologist trained at UCLA and then moved to the University of Michigan. My research has been in gout, so I've worked on the American College of Rheumatology's gout guidelines and I have been on the task force panel for measured development so that is my only conflict of interest.

I have been funded by NIH, American College of Rheumatology and just recently submitted a PCORI grant, so here I am.

DR. PACE: I'm Karen Pace. I'm on NQF staff, another Senior Director.

Ms. Franklin: Angela Franklin, Senior Director for this project.

MS. BURSTIN: And there's Dr. Bagley, can you introduce yourself? You have to push the red button that says speak.

DR. BAGLEY: I'm Carlos Bagley. I'm a neurosurgeon at Duke University and I

was nominated by the North American Spine Society.

MS. BURSTIN: And any disclosures you'd like to share with them?

DR. BAGLEY: No disclosures.

MS. BURSTIN: Great, thank you.

And one more new addition. I can't see your name tag. It's still not -- push the button the button that says speak.

DR. VISCO: Oh, just push it, there we go, sorry.

Chris Visco, I've from Columbia University. I'm a physiatrist and nominated by the AAPM&R which is our academy.

MS. STREETER: Thanks and we also have Kim Templeton on the line. Kim would you like to introduce yourself?

CHAIR TEMPLETON: Hi, thank you. I'm Kim Templeton, an orthopaedic surgeon, Professor of Orthopaedic surgery at the University of Kansas.

I believe I was nominated by the

U.S. Bone and Joint Initiative. I have no disclosures and I apologize for not being able to be there with you for the next couple of days. We had some clinical issues here to take care of.

MS. BURSTIN: All right, thank you everyone.

Steve just joined us as well. Steve Brotman, can you introduce yourself?

DR. BROTMAN: Hi, I'm Steve Brotman from AdvaMed and I have no disclosures.

MS. BURSTIN: All right. So thank you for all those introductions and disclosures. Just one thing we always ask of committees is, you know, you had your opportunity to give your opening disclosures but if at any point during the course of this meeting you have any concerns about potential conflicts or disclosures, please come forward to me or to the chair or any staff and we really want to try to address those issues as

quickly as we can.

So of course, you've already mentioned that you're on these measures for gout, so we will ask you to recuse yourself from that discussion. You don't have to leave the room, but we can ask you not to participate in those discussions, obviously.

And thanks everybody, I'll turn it back over to Katie, I guess.

MS. STREETER: Thank you.

So just to set some ground rules for today's meeting, NQF has been working to improve the committee meetings based on input from a variety of stakeholders. We've made a few changes in our meeting process. Measure developers will briefly introduce their measures as they come up for discussion. They do have two seats at the table.

Selected workgroup representatives will then begin the discussion of the measures in relation to the Measure Evaluation Criteria.

So we'd really ask that as you're walking through each measure, you keep your comments and your discussion specific to that criterion that we're on.

As is the case with committee members, developers may put up their cards to indicate when they wish to respond to comments raised or correct any statements about their measures.

During measure evaluation, committee members often offer suggestions for improvement to the measures. These suggestions can be considered by the developer for future improvements. However, the committee is expected to evaluate and make recommendations based on the measures per the submitted specifications and testing.

Just some more ground rules. We ask that you are prepared, having reviewed the measures beforehand. We ask that you please base your evaluation and recommendations on the NQF Measure Evaluation Criteria.

Please remain engaged in the discussion and attend the meeting at all times.

Please keep your comments concise and focused and avoid dominating the discussion and allow others to contribute.

An overview of the NQF Consensus Development Process, also know as CDP, this is an eight step process for measure endorsement. We are currently in the Standards Review step, which is committee review of submitted and maintenance measures, newly submitted measures. This is when we'll make recommendations for endorsement.

After this meeting, staff will prepare a draft report that summarizes your recommendation. We'll be posting that report for a 30-day public comment period and NQF member period. We ask that you suggest your peers make comments and we will keep you informed of the dates when that's posted. The link to make comments and everything on our

website.

After comment period, the committee will meet to discuss the comments and prepare responses. We'll then have a 15-day NQF member voting period followed by CSAC review. CSAC is the Consensus Standards Approval Committee.

The Board of Directors will then ratify your decision followed by an appeals period, 30-day appeals period.

So this isn the NQF Measure Evaluation Criteria, nothing new to you. These are the conditions for consideration. Importance to measure and report and scientific acceptability are must pass criteria. If measures do not pass these criteria, they will not move forward through the discussion.

If they do pass these criteria, we'll discuss Feasibility and Use and Usability followed by an overall recommendation for endorsement.

And then tomorrow, if we do have any identified harmonization issues, which we may, with two of the measures, we'll have a discussion at the end of the meeting to discuss selection best of class or how we can harmonize those measures.

And now we'll turn it over to Angela to give an overview of our portfolio.

MS. FRANKLIN: Thanks, Katie.

So we'll move on through the review of our Musculoskeletal Portfolio and one of the new pilots that we're testing in this particular project is the standing committee concept.

And for those of you who may have served with NQF before, this is a new concept and we have a new function for the committee and that is to oversee the Musculoskeletal Portfolio.

Your responsibilities will include, as you can see, providing input on the portfolio which we'll step through in a

moment.

Be aware of which measures are in the portfolio and how they fit into the context of the measures before you for review.

Be aware of other NQF measurement activities that relate to musculoskeletal. At this time, we are the only musculoskeletal activity at NQF, but we anticipate, as the committee goes forward, additional activities will become relevant.

Be open to external input on the portfolio. That means public comments that come in during the public comment period, following this meeting as well as future comment periods.

Provide input from your perspectives as to how the portfolio should evolve over time. And that includes providing your input on gaps that you recognize in the portfolio, areas that you recommend for future measure development.

And again, as I mentioned earlier,

consider the measures already in the portfolio when evaluating individual measures.

At this time, I just want to step through the musculoskeletal disease and the definition that we have right now currently for musculoskeletal disease.

And as you can see on the slide, that covers injuries or disorders, including inflammatory and degenerative disorders affecting the muscles, nerves, tendons, joints, cartilage and supporting blood vessels, disorders of nerves, tendons, muscles and supporting structures for the upper and lower limbs; neck, lower back that are caused or participated or exacerbated by sudden exposure to sudden exertion or prolonged exposure to visible factors such as repetition, force, vibration or awkward posture.

Just to give you some context, movement for nearly half of Americans over the age of 18 and many children is restricted by

musculoskeletal disorder that includes arthritis, back pain, fractures, osteoporosis, sports traumas and any other elements that affect mobility and function.

Prevalence of musculoskeletal disease is significant. It's the leading cause of disability in the U.S. and the prevalence of disease requiring medical care has increased by more than two percentage points over the last decade and includes now more than 30 percent of the population.

And this is a graphic that we have sourced from the U.S. Bone and Joint Initiative publication, The Burden of Musculoskeletal Disease in the U.S. And this just graphically shows the increasing prevalence of musculoskeletal diseases in proportion to the total population.

Some additional context, 89.7 million individuals have cited musculoskeletal disease as a primary health concern in response to the Medical Expenditures Panel

Survey that was conducted during 2004 to 2006 and in 2008, the number of adults reporting musculoskeletal disease increased to 110.34 million in the National Health Interview Survey.

So, there's been more than a 47 percent increase in the total aggregate direct cost to treat persons with musculoskeletal disease during the same time and some estimates place annual direct and indirect costs at $287 billion, again, that's from the National Health Interview Survey.

Over the period of 1996 to 2004, the proportion of persons with one or more major subgroups of the musculoskeletal diseases has risen along with arthritis, joint pain, spine conditions being the most prevalent.

And again, I have a graphical representation of these facts. And again, we sourced this from the U.S. Bone and Joint Initiative, Burden of Musculoskeletal Disease.

So I wanted to give you a little history of work so far by NQF in the area of musculoskeletal disease and this is comprised of Consensus Development projects that have been held over the years from 2009 to present beginning with our Outpatient Imaging Efficiency Project in 2009, our Ambulatory Care Standards using clinically enriched data in 2010 and our Ambulatory Care Project which endorsed additional outpatient measures also in 2010.

From all fo these projects, we ended up with 26 endorsed musculoskeletal measures as of 2011. However, for various reasons, most of those measures have since been retired.

So that leaves us with actually four existing measures in the portfolio. Two measures were endorsed in the Clinically Enriched Administrative Data Project and they are the Measure Number 0054 related to arthritis and the use of DMARDs as well as

0052 low back pain use of imaging studies, both developed by NCQA and they are part of the NCQA Back Pain Recognition Program.

The next two measures that we have in the portfolio are Number 662, median time to pain management for long bone fracture. And that was endorsed during our additional outpatient measures that were endorsed in 2010.

And in our Imaging Efficiency Project, the measure number 0514, MRI of lumbar for low back pain which was developed -- both of these measures were developed by CMS.

So as we've defined our portfolio, we are focused on arthritis and related conditions as well as musculoskeletal injuries. You might also think about congenital and developmental conditions, neoplasms of bone and connective tissue and osteoporosis and bone health as well as spinal deformity and related conditions under this

umbrella.

However, those conditions are divided amongst our other portfolios with spinal deformity and related conditions are considered in the Child Health and Material Portfolio. Neoplasms are considered in the Cancer Portfolio and osteoporosis is included in the Endocrine Portfolio.

So that leaves us with our current portfolio up for review during this phase in the areas of arthritis and related conditions as well as musculoskeletal injuries.

Our measures fall into the topic areas of Timely Pain Management, Imaging, Screening and Assessment for Rheumatoid Arthritis as well as Therapy for RA.

We're also considering gout measures in the areas of Assessment, Monitoring and Therapy.

So we have a total of 12 measures for review and in the portfolio, so that will certainly be the topic of discussion as we

move into Day Two.

And again, this is a listing of the particular measures as they fall into the topic areas.

And for the context within the National Quality Strategy, musculoskeletal disorders measures in this portfolio fall within the Safety and Affordable Care domains of the National Quality Strategy which also serves as our north star as we conduct our work today.

So, as a product of our previous projects, prior priority areas for gap filling have been identified in the areas of Disease Modifying Anti-Rheumatic Drugs, or DMARDs, osteoarthritis, care for low back pain including appropriate utilization for low back pain treatments and the management of low pain including manipulative treatments, oral steroid use, narcotic use and functional status outcomes.

So as we go through our portfolio

and, again, we'll get to this more on Day Two, here are your questions to consider.

What are the high leveraged improvement opportunities in this area?

Why are the measures current in the portfolio important?

And do the measures adequately address quality issues?

And also, primarily consider other areas of musculoskeletal disease and disorders and whether you're aware of any measures or concepts that should be brought forward for consideration or considered by a measure developers for future development?

MS. BURSTIN: And just one thing quickly to add to that, we really are very interested in finding out measures already in use and there are places where you've got measures that we are using now that we would like to increasingly look at the prospects for those measures and try to bring them in and make them into national standards rather than

them starting from scratch in that very long development cycle.

So especially because this is a standing committee, we'll be with you for a while so we'll continue to ask you to recommend even measures, for example, used in research. Many of you have grants for NIH or hopefully PCORI.

You know, as some of those measures potentially are very useful in research, some of those might be things we could apply and bring forward for national standards as well since, as you can see, there are a lot of gaps in this area.

MS. FRANKLIN: Thanks, Helen.

And that will move us to our consideration that's key for today's measures and that's our eMeausre Trial Implementation Pathway which we are piloting in this project. And there are several measures within the project that could fall into the pilot.

Most notably, we have the four

gout measures which are also eMeasures that will fall into this pathway, as well as a rheumatoid arthritis measure that will be a potential candidate for this pathway.

And Karen, I don't know if you wanted to speak a little bit to the pathway and how we conceived it or is we want the trial measure

DR. PACE: Right, so basically the reason for this pilot eMeasure Trial Implementation, and I think we're calling it Trial Measures.

But the reason for it is with eMeasures, given the current uptake of EHRs inability to embed eMeasures and find testing sites, that that's been a limiting factor of bringing eMeasures to NQF.

There's not enough sites to implement the eMeasure even for limited testing. And so, in an effort to continue progress in this area, the thinking was that we would look at a process for approving

eMeasures as Trial Measures so these would not be considered endorsed but they would essentially have to meet all of the criteria except the scientific acceptability formal testing for reliability and validity.

So it has to already be completely specified as an eMeasure, meaning the HQMF, using the HQMF standards for specifying an eMeasure. They would have had to have done the eMeasure feasibility assessment and that includes kind of the data element feasibility as well as testing once this project actually runs.

So the idea is that these eMeasures are ready to implement and by giving it approval as a Trial Measure, hopefully can facilitate the testing that would then occur getting enough sites to participate in the testing and since they would data to actually then come back with formal reliability measures.

So that's the issues and we're

going to be pilot testing it I mean with some measures in this current project.

MS. FRANKLIN: Perfect, thank you, Karen.

So with that, questions? Yes, Mark?

DR. JARRETT: On the eMeasures, would you consider, because I know on one of them there's an eMeasure and there's the same thing in a non-eMeasure. Would we run parallel because that would be a really good test because on the abstracting it the old-fashioned way while electronically trying to measure it to look for reliability and validity?

DR. PACE: Yes, that's a possibility. I don't know, are they different measure developers?

MS. FRANKLIN: No, they're all the same.

DR. JARRETT: They're the same.

DR. PACE: Interesting. So,

generally that is how eMeasures are being tested right now is to look at the data that is generated through applying the eMeasure compared to record abstraction, you know, using the full record to abstract the same elements.

So there's definitely some possibility, especially with it being the same measure developer.

Go ahead, yes?

DR. JARRETT: No, just because there's a warning and I imagine some of the other people in the room have the same thing as we look through meaningful use and we start paralleling our eMeasures that we send in for meaningful use with what we do by abstraction.

The numbers we sent to CMS are very different and some of us get very nervous sending CMS two sets of data because, even though they claim and they sent us a letter saying it's not a problem, we worry about that.

So it we have to keep that into consideration why you may want to pilot it in selective places.

MS. BURSTIN: That's exactly what we're actually hoping to learn more from. I actually am chair of the Quality Measures Workgroup for the Health IT Policy Committee, so it is in fact true that the measures you're sending in for meaningful use are simply based on whether you could report not the levels or performance.

But nonetheless, we all want to understand what those differences are between tarp-based measures, claims-based measures, other approaches and eMeasures. So lots to learn there. Both ambulatory as well as in the hospital.

MS. FRANKLIN: Any other questions about anything we've discussed so far?

Okay. So that moves us to consideration of our measures and, as I mentioned earlier, we will start with the gout

measures and the developer for those measures are the American College of Rheumatology and we do have two seats at the table for representatives from ACO.

MS. STREETER: Actually, I think this is a good time to pause and go over who we'll do the voting using the clickers that you have there.

Ann will walk us through that. We can even do a practice run to make sure that everyones clicker is working and your vote is registered.

MS. PHILLIPS: So, I'll explain to you how the vote stamped devices work. I think all of you should have a fob that I've passed out to you. You'll need to point your votes snap device at -- it works on line of sight at this computer. There is a USB dongle, so just point it here.

It works on line of sight. You should see a red light on your device. We give you 60 seconds for voting. The device

only records you last input. So if you vote and want to change your input, just press the next button. There's no need to clear it, just press the next button. But it has to be done in that period.

So we can go ahead and try a sample vote right here. Okay.

DR. PACE: So the voting slides are on the two screens at the end of the room.

MS. PHILLIPS: I don't know why this just crashed. Hang on just one second.

DR. PACE: So while we're making sure the program gets up and running, I just wanted to -- I know you've been through this on some of your calls, but again, as we talked about at the beginning, we really want you to follow the criteria in terms of your recommendations and you've seen these before but they are at your seat.

The algorithms for going through the evidence, the clinical evidence criterion and also the reliability and validity.

So, I'll just make a few comments about this and then we can see if you have any questions because I know you've all been reviewing measures and probably have encountered some questions as you've been going through those and thinking about how they meet or don't meet our criteria.

So, I think, are most of these measures process measures? So we have a different way that we look at outcome measures but most of these measures will be process or structure measures, so we really do want to look at the strength of the clinical evidence for the focus in that performance measure.

So, the idea is direct evidence and also what's the quality, quantity and consistency of that body of evidence? One of the things that we want to distinguish is that clinical practice guidelines may be the source of a systematic review of the evidence where the quantity, quality, and consistency of the body of evidence has been reviewed.

Not all guidelines are equal and, you know, so if it's, you know, we have on the submission form where we've really asked the developers to indicate what was the source of the systematic review of the evidence to provide information about the grading of that evidence and actually to provide the summary of the systematic review of the evidence.

So one thing to keep in mind is if it's based on a guideline and there is no summary of the quantity, quality and consistency of the evidence, that the highest possible rating would be moderate and that's only if the grading and evidence really indicate that it's fairly strong evidence.

Guidelines that are basically recommendations that doesn't fit into our evidence criteria and those would need to be, if they're considered at all under the exceptions to the evidence.

And in the algorithm at the very end, beginning with Box 10, we have -- kind of

walk through how you might consider and exception to the evidence.

So the first question is, are there or could there be performance measures of a related health outcome or evidence based intermediate clinical outcome or process instead of what's being presented? And then again, is there a systematic assessment of expert opinion that that should be done and does the steering committee agree that it's okay to or beneficial to hold providers accountable for performance in the absence of empirical evidence?

So, the other thing I'll just make a distinction about is this idea of direct evidence. So a lot of times, we may see measures that come in, assess a lab value, assess a label value, assess a blood pressure value. And obviously, you have to do those things, but the evidence is really about, you know, the relationship of the actual blood pressure value to the health outcomes and

mortality/morbidity or there's evidence about treating something.

So there's really not going to be, and we don't expect, that there's going to be body of evidence about the need to assess blood pressure.

But again, that's an example of where we would much prefer to have measures about how you treat blood pressure or what the control fo blood pressure is rather than measures of Just assessing it.

So, just keep that in mind and we can certainly answer questions as you go through this.

The other algorithm is for the reliability and validity and we will, again, we can go through these maybe more specifically as you're going through the first measure if you have questions. But I think tow key things about reliability and validity is that the high rating is reserved only as potentially eligible for measures that have

been tested at the performance measure level of the performance measure score, not the data elements.

So, just a quick example is that if people are looking at the data that are used in the measure, for example, interabstractor reliability, that's at the data element level. What we're talking about when we talk about reliability of the performance score, it's really a signal to noise analysis, being able to distinguish between provider versus within provider differences.

So obviously, the performance scores are what's going to be used in accountability applications so that's why that would be eligible for a high rating than any of the testing and how the performances score.

Again, it's not just that they did the test, but was it an appropriate test, appropriate sample and adequate results.

So, high might be the highest but

maybe the results are in the mid-range and you might want to still rate it as moderate.

One other point about validity, testing is face validity testing would only be eligible for a moderate reading. It's the weakest form of testing and that's where it would fall with the rating scale.

So I'm going to just stop there and see if you have any specific questions about our criteria that I can answer or things that you noticed as you were reviewing measures before we get started and then we can, you know, all obviously work through these things as they come up during your evaluation.

MS. FRANKLIN: We'll be able to demonstrate voting with you shortly, but in the meantime, just to follow on to Karen's comments, we do have a format for going through the measures today that's quite specific regarding how the measures are teed up.

So, as the committee members may be aware, we're going to ask that you follow the process of first letting the developer provide a brief overview of their measure and rationale behind the measure and then we'll ask for the lead discussants to introduce their measure including the measure number, the title, the description and the level of analysis, if that hasn't already been covered by the developer.

And then we'd ask you to walk though each criterion beginning with evidence to discuss the committee's discussion and then throw open the floor to the full steering committee for comments and questions. And we'll proceed in that same fashion through the rest of the criterion and at the end of the discussion of each criterion, we will conduct a committee voting.

So that's the step wise fashion that we'd like to proceed in today.

Dr. Chou will help guide us

through that discussion and facilitate discussion amongst the committee members.

Any questions about that process? Yes?

DR. ANNASWAMY: Could we test the voting clickers?

MS. FRANKLIN: We will be testing the voting clickers. I was just giving them a little time to get that all set up for you. But we'll definitely be conducting a test vote.

Are there questions about the process or anything that was discussed?

So, as we get the voting materials ready, I think we can go ahead and begin discussion of the first measure, voting by that time, we should be able to have a test vote and then an actual vote following discussions of criterion. So, yes.

So our first measure for discussion today is Measure Number 2549 entitled Gout Serum Urate Target and our

measure steward who is at th at the table is the American College of Rheumatology. Our lead discussants are James Daniel and Steven Brotman who will tee off after the developer.

DR. YAZDANY: And so, Katie, can I just ask you to bring up the introductory slides?

I'll go ahead and get started here and I'll ask you to advance the slides since you have the controls over there. Yes, yes.

So I am Jinoos Yazdany. I am an Associate Professor of Medicine at UCSF and we're practicing rheumatologists and a health services researcher. I have co-chaired the ACR's Quality Measures Subcommittee. I am the principle investigator on the Rheumatoid Arthritis Measures Project.

My research funding comes from the NIH NIAMS as well as from PCORI. I have no other financial disclosures. In particular, I have no financial relationships with any entity that makes products or provides

services for rheumatoid arthritis or anything else in rheumatology.

I am joined here by Dr. John Fitzgerald who is Associate Professor of Medicine at UCLA. He is the Chief of the Division of Rheumatology at UCLA. He chairs the ACR Guidelines Committee.

He's the principle investigator on our Gout Measures Project. And John also has no financial disclosures.

John and I are volunteers, we are not employed or we are not paid by the American College of Rheumatology.

We're also joined here today, I'll just introduce Rachel Myslinski who is our VP of Quality and Registries as well as Melissa Francisco who is Senior Staff.

Next slide?

So I thought what I would do today is just in a few minutes provide an overview of the methods to provide context for how we arrived at this point.

The very first thing and this was, you know, many years ago, was that as a professional society, we actually set priorities for Quality Measure Development and that was through the development of a white paper on quality measurement which was published in 2010. And through that process, we came up with rheumatoid arthritis and gout as top priorities.

And the reason for that was because of the prevalence, the perceived gaps in quality and the fact that we felt that measures would have a big impact.

We then went on to define methods, and our methods Actually begin with the writing of guidelines. And so both of the measured projects that are being presented today happened after the development of a guideline in rheumatoid arthritis that was in 2012 and in gout in 2013 and workgroups that were multidisciplinary in nature with very strict conflict of interest policies were

assembled.

Next slide?

The workgroups defined the measure concepts and really, we tried to build on the lessons that we'd learned from quality measurement over the last decade. And the central task was how to push measurement forward to decide what's meaningful and figure out how the measure that.

After those initial measure concepts were developed, we assembled expert panels and this was a multistakeholder participatory effort where we had practicing rheumatologists. We solicited a nominations from other professional societies.

So for example, the American Academy of Orthopaedic Surgery, the American College of Physicians and others were involved. There were patient pairs, so for example, a Medical director for Medicaid Managed Care Plan was on our panels, rheumatologists and also allied health

professionals.

We used the RAND/UCLA Appropriateness Method for ratings and provided the panelists with evidence reviews, pre-conference anonymous ratings and then post-conference ratings.

Next slide?

Measures that passed the expert panel then went on to a public comment period and the ACR represents 90 percent of U.S. rheumatologists and every member gets an email through the Rheumatology Morning Wire as we call it, every morning so through that and also through newsletters and other efforts, we asked for public comment on the measures and we also targeted a request for responses from other stakeholders, including other special societies and especially societies and patients.

And then finally, the various levels of leadership at the American College of Rheumatology approved the measures.

So as you can you see, this was a very intensive consensus process arriving at this point.

Finally, we got to the -- for the rheumatoid arthritis measures, the e-specifications a well as the recruitment of testing sites to do the feasibility testing, the query building and the validity testing.

And I just want to take a moment, actually if you could advance the slide, just to thank the AMA PCPI who were instrumental in helping us navigate the testing since we are the new measure developers and this process is quite intensive, so it was nice to have their expertise and also the NQF staff who provided feedback from the measures concept audit.

I'm almost done here, next slide.

This is a very brief overview of rheumatoid arthritis.

About 1.3 million American have RA. It's more common in women. It's a chronic disease that has no cure and without

treatment, one-third of patients will have permanent disability in five years. And as rheumatologists before the modern era where we have so many medications to pick from, it was not uncommon for our waiting rooms to be filled with lots of assistive devices and even wheelchairs.

And with modern treatment, most patients can expect that their pain will be well controlled and this joint damage will be prevented.

Next slide?

You may be wondering why rheumatoid arthritis is a top-20 Medicare prioritized condition by the National Priorities Partnership. Well, I think that if you're CMS, you see that the main age in this disease is 67. This is a chronic disease and so with the prevalence is actually going to increase in our Medicare population as the population ages.

We have good treatments and so

providing effective care is obviously a priority and I think we would be dishonest if we didn't acknowledge that fact that the cost of treatment is very high and just to illustrate this point, I made this table with the top ten best selling drugs in 2013 by sales and revenue. So these are the ten in the entire healthcare system and four of them are in the area of rheumatoid arthritis.

So, you know, there have been lots of debates about Hepatitis C and the cost of those medications recently, but we've had to struggle with drug costs in RA for a really long time.

Next slide?

The past decade, really the primary RA quality measure has been the use of DMARDs. I call this an equity measure because it has really been an area where we've identified the severities and monitoring has increased use.

But new measures that we're

proposing today are attempting to build a measure and an infrastructure that's patient centered and aligns with the National Quality Strategy to enable the measurement infrastructure will allow us to look at effectiveness, safety, population health and eventually value.

And our measures address three main areas. The first is what is the most important outcome to patients? That's the first measure, functional status.

What's the most important outcomes to clinicians?

What is it that we base our clinical decisions on?

What are the outcomes of clinical trials? That's disease activity.

And what's the most critical thing for patients safety that we can measure? And that's TB screening.

Next slide?

Gout. Gout is interesting because

the prevalence is increasing and we think that this has something to do with the epidemic of obesity in this country and a recent NHANES survey, a remarkable 3.9 percent of adults reported gout.

I think sometimes we underestimate how severely gout can affect people. It's associated with excruciating pain, it can destroy joints and lead to disability. And as rheumatologists, we see widespread quality and safety problems in this disease more than any other. And so that's why this was a priority.

And our gout quality measures -- next slide, this is my last slide -- address these questions.

What are the cornerstones or appropriate treatment? That's urate-lowering therapy and prophylatic therapy.

What is an important intermediate outcome? That's the uric acid target.

And how should drugs be dosed? Uris monitoring.

So with that, I will turn it over to Dr. Fitzgerald. To introduce the very first measure.

DR. FITZGERALD: Thank you, Jinoos.

What I'd like to do is, I've prepared a handout for the group that was distributed focusing on responses to the questions that were addressed at the tele-conference, tele-meeting two weeks ago.

MS. FRANKLIN: So, first, if you could just introduce the measure that we have coming up next and then as the committee discuses the measure and those questions are raised, we will be happy to have you respond.

DR. FITZGERALD: Okay. So, my understanding is we'd have an intro that we can give an overview before the discussion.

MS. FRANKLIN: Yes, of this particular measure.

DR. FITZGERALD: Okay. So for the first measure which is 2550, the title is Gout

ULT Therapy, again prepared by the American College of Rheumatology and a brief description of the measure.

It's a measure of the percentage of patients age 18 and older with the diagnosis of gout and either and either tophus or tophi or at least two gout flares or attacks in the past year who have a serum urate level greater than six milligrams per deciliter who are then prescribed a urate-lowering therapy.

And so, in introducing this, I wanted to use this measure as a chance to introduce some of the rationale for the gout measures in general and as RA was described as a disease that there's no current cure, gout has often been called a curable disease, yet we all see patients with advanced tophaceous gout on occurrence and attacks and the reason for this is there's a lot of gaps in the quality of care. And some of those gaps are highlighted on Page 2 of the handout.

One of the notable problems is that adherence to urate-lowering therapy is quite poor. In several studies, it's been documented that adherent rates of allopurinol for up to 30 percent of the population revealed that patients are only taking their allopurinol ten percent of the time as prescribed.

Other documented studies by Halpern and Colleagues have shown that 50 percent of patients are described as nonadherent meaning a medical possession ratio of less than 80 percent.

In those patients who are nonadherent there are higher serum uric acid levels and in those higher serum uric acid subgroups, there are higher rates of gout attacks.

In addition to poor adherence, it's also been noted that there's very little titration of urate-lowering therapy. Patients are often prescribed a single dose. I can

cite three studies here where 300 milligrams or less are often initially prescribed and there's no titration that follows up.

Furthermore, there's no serum urate monitoring or there's very poor serum urate monitoring done. In studies anywhere between ten or 20 percent of patients might have a serum urate check in the following year.

This is problematic because Perez Ruiz and colleagues have noted have noted that most patients require a dose of over 300 milligrams to achieve the serum urate target of six.

In a study by Annemans and Colleagues, it was a large study looking at 4,000 UK patients and 3,000 patients in Germany, there was a greater adherence in the UK with better serum urate control and fewer flares, but again, poor adherence in both countries was noted.

I'd like to address some of the

concerns about serum urate here at this time as well. I wanted to point out with the charts that are on Page 4 is that the prevalence of gout and the instance of gout is rising over the years but despite that, there has been no change in the prescription rates and the treatment of patients with hyperuricemia gout and those have remained at 20 percent.

So on the following page, another author has proposed a quality model, a conceptual model, this is by Dr. Lin and Colleagues and there are many areas that could be addressed in the management of gout. We're not able to address those areas.

We focused on the physician-patient provider intervention and I want to highlight a couple of things on this chart.

On the left side of the chart in the system description of things that could be improved for gout, one thing that these doctors noted is that gout doesn't have any

quality indicators and we're hopefully going to address that here.

In the areas that have been described for improvement, it's under use of urate-lowering therapy and unsure knowledge about the indications for urate-lowering therapy, the lack of prophylaxis, unaware or lack of knowledge about the target and duration of therapy and lack of dose escalation.

The next point that I'd like to address is there were a lot of concerns about the validity of some serum urate as a marker. And two -- actually, I'll table that until we get to the discussion on that.

So with that, we'll open it up for discussion.

MS. FRANKLIN: Thanks Dr. Fitzgerald. So that moves us to our lead discussants for the measures and that's Drs. Daniel and Dodge, if you could give your overview of the measure from the workgroup

perspective and then we'll -- yes, go ahead.

DR. DANIELS: Were we supposed to do 2549 or 2550? Because he just --

MS. FRANKLIN: 2550, I'm sorry.

DR. DANIELS: Okay.

MS. FRANKLIN: 2550.

And just a reminder that these are Measures that are candidate for the Trial Measure Approval that we discussed earlier.

DR. PACE: So, we're going to go through the subcriteria evidence first or --

MS. FRANKLIN: Yes, we'll be going through -- we discussed the procedures and measures and then we'll start with the main criteria evidence, importance to measure -- of importance to measure and report and then we'll be voting, but we will conduct a test vote prior to your actual vote.

DR. DANIELS: I just want to use people on the panel here and I don't know if I'm the best person to start this, but I'll give it a shot.

So am I supposed to review the element of the workgroup first and then I read the --

MS. FRANKLIN: Yes, if you could just, yes, give an overview of the measure, the description, a quick description which we have and then the workgroup summary regarding evidence.

DR. DANIELS: Okay, a brief discussion was percentage of adults greater than 18 years with a gout diagnosis, being prescribed a urate-lowering therapy and there was a lot of discussion right at the beginning of the measure on the 1(a) evidence report as the measure.

And the first, I won't read it verbatim, but, there was comments made that they felt that people felt that there probably were some more information but it wasn't presented, and it looks like more that was done.

And then there were a number of

questions around how these would attacks would define measurement of -- and they also felt that with the performance gap that there was some evidence of under treatment in this population and they felt that it would have an impact, if it was proven that this was there.

And the workgroup agreed that the evidence presented did not directly support the measure focus and the study showed that between patients on a urate-lowering therapy reduces the number of attacks should be cited and strengthen the rationale for measure.

There was also some questions about clarification on why the serum levels was chosen, how that happened.

MS. STREETER: I just wanted to jump in and ask everyone to speak really close directly into your mic. We've been asked by our court reporter, he's having some problems picking up the voices. Thank you.

CHAIR CHOU: Does Dr. Dodge have any additional comments?

DR. DODGE: No additional comments.

CHAIR CHOU: So maybe we'll open it up now for questions or discussion from the panel. I believe you have something you wanted to say first?

DR. FITZGERALD: I can respond to those questions if you'd like at this time or wait.

CHAIR CHOU: That would be great.

DR. FITZGERALD: Okay.

So we responded to the request for additional data on Page 6 of the handout. We have cited the three studies -- no, not that one. There's a Word document, roughly 12-page. The first page is a summary of the measures and if you go past the fish on Page 6 to the 2550 Gout ULT Therapy Heading.

And what I've done is posted the questions that were addressed to us and so one of the questions was about evidence and we're providing the citations for the three studies

looking at febuxostat, these were two randomized controlled trials and an open label follow-up trial demonstrating that febuxostat lowered serum uric acid and reduced the frequency of gout attacks.

Also provided here are articles documenting the efficacy about allopurinol. Not surprisingly, these are older articles dating back to the 1960s, but they describe the effect of allopurinol on lowering both uric acid and frequency of attack and tophus reduction.

There's a single article also cited for the use of probenecid and then two articles that are cited for the use of pegloticase and the intravenous uricase that's used specifically for patients with advanced tophus.

Regarding questions about attack definition, that's been updated. I don't have that on my copy but Melissa provided updates for that.

We are going into, this is the eMeasure for the proposal for the trial testing and we are going into testing. We are proposing to use ICD-10 coding and so we'll be looking at that during the specification phase of our study.

The performance gap was acknowledged and the last question was about the serum uric acid as a level of six milligrams per deciliter. I do have some more data on that but briefly, the six milligrams per deciliter has been recommended and endorsed by the British Society of Rheumatologists going back to 2004. EULAR, the European League Against Rheumatism and their guidelines in the 2006 and again, by the American College of Rheumatology.

The rationale for a level of six is that in an article by Soji and Colleagues --

CHAIR CHOU: Can I stop you for a second? I think we'll wait to talk about the

uric acid targets until we get to those measures then we can focus on the treatment piece first, if that's okay.

Do people have other comments or questions to ask about the supporting evidence?

DR. DANIELS: If you want, I don't know, I didn't get this until last night.

But I did go through those so if anybody would like a summary I can help with that.

CHAIR CHOU: That would be great.

DR. DANIELS: Okay. What they did was, at least to me -- what they did is I thought they did a good job on clarifying what they were measuring. It was a little more clear and what they basically said to be in the study, you have to either have typhus or tophi. You have to have two gout flares in the last year and you have to have erosions on a radiograph.

And the way the presented seemed

at least to me to clarify things a bit better.

The big study has been brought on by Halpern, it was like almost 2,500 patients and what it really studied was if the patient might take the allopurinol, it really didn't talk about it, it was prescribed, it's just that the people who they prescribed didn't take it.

The second one by survey basically looked at medication possession rate, so of the 30 percent of patients that had medication, less than 10 percent kind of were still using it. So it looks like that there are any of them prescribed and not given.

And then the last one talked about patients who have received medication and it was only 300 milligrams, but the dose rate was below six was only 370, excuse me, it was 370 is what they said the average to bring it up, so it wasn't that far off.

The Japanese guidelines, you know, were kind of mentioned and had gotten sent out

in an email and that had a summary where they actually used their definition was seven milligrams and that kind of goes along with I think with 2011 some of the later guidelines on that.

And then we've had a real big study in the general practice research database which eight percent of the UK population were followed. And out of those, there was really no change in proportion to the people that were adherent.

The fishbone diagram, I thought really brought out a lot of the good points that they were trying to bring out. And then they gave a lot of their studies talked about the latest ones with the -- I'm going to mispronounce this, with the febuxostat.

And the allopurinol was surprisingly how low studies were. They were all old and they were like 106 patients and the average was like 30, 33 and 12.

And with the probenecid, there were only

82 patients.

When they talk about feasability, they basically referred to the other guidelines that had been used and the quality and care of patients.

CHAIR CHOU: Are there questions?

I had a couple.

So one is that at least in the febuxostat trials, you know, you have to have pretty high uric acid levels to get into the study, I think over eight, and then you know, a lot of the other studies focus on people with pretty advanced gout.

And so, I guess one of the questions is, you know, what is the evidence in people with minor or less severe attacks of gout?

And the other related issue is something that came up on the call which is about, you know a lot of, in primary care, frankly, a lot of people are kind of given a clinical diagnosis of gout and it's often not

confirmed. So with, you know, a joint fluid analysis and a lot of them are pretty soft calls.

And, you know, somebody says my foot hurts and they get treated like they've got gout. And I guess the concern is about possible potential over treatment is that -- or some people -- with some people getting treated who don't really have gout. We've given a lot of these patient, I don't think that -- I mean the guys you see are, you know, different than the ones that are seen in a lot of primary cares.

So I guess two questions about kind of the populations studied in the trials and then kind of the reliability of the diagnosis, particularly in primary care settings.

DR. FITZGERALD: Okay, thank you. I'll take those in order if you don't mind.

So I believe the first question was about the Japanese study. The Japanese

guidelines are interesting, they're nicely written. In Japan, they treat asymptomatic hypertyrosinemia and so part of the Japanese guideline measure was to develop the spectrum of patients with gout through patients where they had asymptomatic hyperuricemia.

And they developed a novel nomogram looking at what they defined as a normal serum urate as less than seven and hyperuricemia as greater than seven. And so that became their threshold.

It has been argued by many about which threshold is used. The problem I have with the Japanese guidelines is that they recommend treatment of patients with hyperuricemia who don't have gout and that's not something that's been accepted here as showing a clear benefit.

So, I wouldn't focus on the seven milligram definition from the Japanese guidelines other than to say that there is some variability out there but the three major

--

Actually, I'll even go back and add the Dutch guidelines which were the first ones that are out. But the four major guidelines that have focused on this have used six milligrams per deciliter and the Europeans have even used five milligrams per deciliter in the more severe cases as the DACR guidelines.

For the second question about the less severe patients, one of the -- part of the specification is that this is trying to capture patients who have more severe disease by those patients who are having frequent gout attacks, two or more per year. Those who have tophi or those that have a erosions.

For the patient who's having an attack every 14 months or even just one attack per year, they wouldn't be included in these guidelines.

So I think the less sever patients would be -- I mean they would not be included

in this measure and, you know, hopefully then that wouldn't create the concern about over treating the rest of your patients.

And finally, regarding the clinical diagnosis of gout. Gout can be tough to diagnose even in rheumatology practices, the majority of the patients were not needle confirmed cases.

In fact, if you look at the criteria, several of the criteria have sprung up over the years to get around the needle confirmation. And now there are clinical indications for the diagnosis of gout which have high likelihood ratios in the 10 to 30 range when you're looking at them.

And so, criteria such as podagra, rapid response to therapy, limited duration are all included in making a clinical diagnosis of gout.

And again, because we're looking for patients who are more severe, not the subtle gout, I hope that and think that a lot

of that would be excluded.

CHAIR CHOU: Yes, I think that's helpful. I mean I guess I still, you know the tophi and erosions are fairly easy when somebody comes in with obvious podagra it's, you know, fairly straightforward.

But I guess my concern are these, you know, I mean I'm a primary care doc, so I see patients who come in and they've been given a diagnosis of gout and I often am not sure that they have gout. They may come in twice a year saying, you know my foot hurts and I --

You know, sometimes I'll even just ask, you know, when we kind of do the history, we'll talk about their gout and they'll say, oh yes, I had a few flares and I took a little bit of Naproxen.

I have no idea what that means or how reliable that is, is that really a flare? That's really what I'm talking about. And are we really saying that those are patients who

we should put on uric acid-lowering therapy? I guess.

So anyway, is there a question?

DR. MATUSZAK: Yes, I just had a question for the developers on this and maybe I'm misinterpreting this a little bit, but did you say in the evidence supporting this, only about ten percent of patients are actually taking the prescribed meds?

So is the outcome that we're going -- I mean is the bar that we're setting just prescribing meds or is it actually that we would like to see the quality outcome being actually driving down the uric acid levels in patients regardless of what method it takes.

I mean, is that kind of like going back to some of the analogies we've used saying that we're, you know, going through -- we're not going to target hemoglobin A1c for diabetes, we're going to target whether or not you prescribe an insulin for it.

So, is this really what you think

is going to be an effective target for getting the quality outcome we're looking for?

DR. FITZGERALD: I think you nicely set the stage for some of the follow-up measures but in the follow-up measures, we do specify the importance of trying to reach and to measure uric acid and to reach a uric acid target.

Uric acid levels and targets are well correlated with patient outcomes. But the first step is getting patients on therapy.

The British study showed that patients aren't on it and then patients aren't taking it but the way you would encourage adherence is by trying to get patients to a uric target.

DR. MATUSZAK: Yes, it certainly seems that that probably as the bigger obstacle and barrier to overcome and I think how you position this one, this particular measure in the prescribing of the med being the process that we're looking at, again,

looking at the available evidence that's out there I think that really the evidence is driving at lowering the uric acid level, not necessarily just by making the prescription, actually getting patients to take it, actually getting them to do the dietary modifications and everything else seems to be a much bigger yield for the quality outcome.

DR. FITZGERALD: I think this measure by itself wouldn't address all those issues. There is interesting evidence on adherence in how the drugs are prescribed or what gets patients to they to stay on their drugs.

CHAIR CHOU: I see a question over there.

I Just wanted to make one follow-up comment first.

I think in order to improve adherence, you first have to get people to start prescribing the med, I think that's generally where we start with the quality

measures.

The second point is that this is the same for many chronic conditions. I mean hypertension, management of hyperlipidemia, management of diabetes, it's less than 50 percent adherence rate so this is not unique to this condition.

Any comments here or questions here?

MS. DAVIS: I was just thinking about what Jason was saying is why don't we just go to the end point here and figure out what the levels, you know, measure the levels.

I'm representing the consumer and the purchasers perspective and maybe one way of getting there is by actually publishing or reviewing the levels of different practitioners, comparing them to each other rather than going through -- letting them figure out how to get the patients to keep compliant, letting them figure out how to change their behavior.

CHAIR CHOU: Are there any other questions or comments?

DR. JARRETT: This is a related question. Disclosure, I'm not a rheumatologist, but could you tell us a little bit about why the compliance is low? Is there a financial barrier to the patients? Is there cause and side effect barrier for patients? Why are patients who are prescribed these medications not taking them?

DR. FITZGERALD: So when you look at the cited evidence, and I have that later in one of our discussions, education is a big component. It's been postulated that not checking levels and not giving patients feedback is another component.

When urate-lowering therapy is prescribed, there's an expected increase rate of gout attacks during the first three to six months. That has a potential negative feedback and so if patients aren't educated about that and if patients aren't prophylaxed

against those attacks, there is a risk of loss of adherence.

And so those are the reasons cited with lack of adherence and the flare rate during the first three to six months was one of the strongest associations. It had a twofold increase. Those patients who flared were less likely to stay on their drugs.

So there's important education and then we think of prophylaxis is important as well.

DR. JARRETT: Is there a financial barrier?

DR. FITZGERALD: Well, there shouldn't be. You know, allopurinol is dirt cheap. It's been around forever. Coltrazine used to be dirt cheap but no longer the case, hopefully again.

The newer drugs are quite expensive. We're really not interested in the big loaded case. We're not addressing that with any of these patients. We're not looking

at the severe or refractory cases really, but the cost of allopurinol, it doesn't cost a lot.

But again, the asymptomatic period between attacks, it's very much like hypertension. You know, RA, there's a good feedback, if you don't take your drugs you hurt. And with gout, if you're feeling well, a lot of patients want to try and come off their drugs and they just don't feel they need anything.

The problem is that those urate crystals, as long as the levels are above six, they're going to start forming and those crystals are auto-inflammatory and can cause erosion and damage.

DR. CHOU: Yes?

DR. DANIELS: Just a follow-up that just from a primary care standpoint.

What happens is most of the time people go to the doctor when they have a problem and they want that fixed and then

otherwise, getting them to come in is a little bit tough.

So there's a lot of confusion because there's a bunch of issues here that are sort of all intertwined and so, you know, they really care, I don't want my toe to hurt. Okay? And so then the doc gives them, you know, an anti-inflammatory which really has no effect on the long-term deposits of tophi and all that to happen.

And I also know we're discussing other issues here besides the level of the uric acid because there are people that come in with attacks of low and often people with higher acid levels, that don't have gout attacks but they end up having damage to other organ parts.

So and then when it comes into it, there's even, I think among providers, a question on, you know, how do you actually diagnose gout? Because for along time, it was like you've got to stick a needle in the joint

and those who have in a primary care setting, you don't get much and then when you get the fluid, that's very difficult, if you're not in the right center to identify. And so they're kind of hesitant to stick a needle in someone's toe and then, you know, what do you do? So there's a lot of guessing.

And so if you're a doctor, you're saying I'm not really sure about this and they come in every so many years because their toe hurts, we probably know that's going to get measured. I'm not saying that's right, I'm just saying that's kind of how it works.

(Whereupon, the above-entitled matter went off the record briefly at 10:00 a.m.)

However, worrying about quality over a large scope of physicians, I worry about the measure because we are measuring something that, quite frankly, we haven't done the education to really teach people how to respond.

And although it is good to say that, all right, this is where our baseline is, and it would provide baseline, and maybe as a trial that is good, I worry about the fact many of the primary care physicians out there do not know how to deal with the first six months when the flareups occur. They don't know how to deal with which drug to use and how to titrate the drug.

And although these measures are good to serve as baseline, I worry that we will spend three-four years seeing terrible results, seeing if the results are valid, et cetera, and that it can't go without a concomitant program. Otherwise, we are just measuring the fact that we don't do well. And I know that we all know that in the general community we don't do well.

DR. FITZGERALD: I am not sure how to solve the education problem. There has been over a decade with four agencies putting out guideline recommendations.

As fantastic as I think the ACR guidelines were, in all honesty, they are not that novel. They really restate what has been stated for years. And yet, the ball hasn't moved on that.

I think there is some lack of appreciation of gout and what it can do. We may have to indirectly thank some of the pharmaceutical companies for promoting gout. It has become more -- you see it on commercials, the guy walking around with that big beaker of green juice. And some of that is motivating patients to ask more about gout.

So, gout really hasn't been on the radar screen. It has sort of been thought you can treat it intermittently. But we do see the patients that advance to damage, erosion, tophi. They undergo surgeries. There is significant morbidity, and the drugs can be quite toxic, as you advance with renal disease and NSAIDs and steroids. And so, it is important to try to treat these earlier.

And a greater public awareness is needed. I am not sure that we are going to have a public campaign.

CO-CHAIR CHOU: I think there is a comment there, and then, Carlos.

MEMBER ANNASWAMY: Has ULT been defined? What are the drugs that count as ULT?

DR. FITZGERALD: So, it would be allopurinol, febuxostat, and probenecid, which are the available agents right now. We are not including pegloticase.

MEMBER BAGLEY: I guess the question I had was, how are the number of flares defined? Is it, as you were alluding to, just what the patient reports or is it that the patient actually has sought medical attention to it? Because I think that is somewhat of a concern for me.

DR. FITZGERALD: Yes, we will be refining that in the specification phase. But, with the ICD-10 coding, there is a code

for gout flares. So, it would probably be, our measure would probably be low on sensitive, meaning that some patients would be sliding by, but the specificity should be good.

CO-CHAIR CHOU: Yes, I think we will talk about that some more when we get to the specifications.

So, maybe are there other, one or two more comments?

MEMBER GHOGAWALA: Just one last question. What number of patients are we talking about here? This sounds like a more severe segment of the people who have two flares a year of tophi. Could you give us some perspective of this population that we aim to study with this measure?

DR. FITZGERALD: I don't know the exact numbers for that, but, of all the gout patients -- and gout patients, again, you saw the 4 percent; it is 8.4 million people. The cocktail napkin calculation is I had 20

percent and Dr. Gardner had 30 percent. So, 20 to 30 percent of patients might be that more severe form. Most patients are really going to be the milder for. That model is difficult.

CO-CHAIR CHOU: Yes, one more comment.

MEMBER SCHUNA: Yes, my question was with regard to it looks like you are recommending urate-lowering therapies, maybe not addressing appropriate therapies in some cases.

For example, a patient with tophi or renal dysfunction or a history of kidney stones probably shouldn't be on probenecid. Is there any cost for that or is an efficiency?

DR. FITZGERALD: I think probably for future development we were trying to keep it simple without going to too many specifications.

CO-CHAIR CHOU: I have a quick

process question before we move on. So, we said earlier that we want to vote on the measures as they are presented.

But, if, for example, there was a motion to limit this to people with tophi erosions or something like that, does that mean that this has to kind of go through the whole process again or is it still possible for it to move forward with some modification?

MS. FRANKLIN: We generally have to look at the measures as they are presented to us. If there were evidence that the developer could present that that modification would be supported within the measure within the timeframe of this project --

DR. FITZGERALD: Our measure is specified for erosions, tophi, or frequent attacks. This is just the summary description.

MEMBER MATUSZAK: Actually, it is in the denominator. When they discuss the denominator later on, that is where they

actually said that --

CO-CHAIR CHOU: Right. Well, what I was referring to was, could the denominator be modified.

MEMBER MATUSZAK: Can I just ask one more question on the evidence before we move on?

I just noticed that the articles here that you cited in this additional piece that you gave us today, this lead author on all three of those. Just so I am clear on this, does that author have any conflicts or any disclosures or anything that might be relevant to the discussion at hand on the evidence?

DR. FITZGERALD: I don't know Dr. Becker's conflicts, but if he was the lead author on all the febuxostat trials, I am sure he received some benefit from authorship.

CO-CHAIR CHOU: Maybe we will have Dr. Daniels kind of look at the evidence and, then, see if we are ready for a motion or a

vote.

DR. PACE: To just do a clarification, from what I have looked at in the submission form, this is an example of where it is based on a guideline. This isn't a systematic review, you know, of a rating review, and consistency wasn't provided. So, we would start at the bottom rating.

CO-CHAIR CHOU: Yes, are you saying we start at the moderate and, then, what has been presented additionally since then?

MEMBER ANNASWAMY: Are we using moderate as the highest it can be?

DR. PACE: Without a systematic review that has been systematically reviewed and rated. So, there is no doubt that there some evidence, that the evidence is here, but there hasn't, at least from what I have seen, there isn't the systematic review where the studies have been graded on the quality of the studies and, then, gone through the whole

process of quantity, quality, consistency.

MEMBER ANNASWAMY: My question was, you said you can start at moderate. I was wondering, is moderate the ceiling or the floor?

DR. PACE: The ceiling.

MEMBER ANNASWAMY: Okay.

MEMBER BUTLER: So, just for clarification, we are walking our way through this algorithm, correct, and we are basing it on this second block, No. 3. And we are saying no to that on the evidence?

DR. PACE: Correct. Well, we are saying -- yes, right.

DR. FITZGERALD: Could I just ask for a point of clarification? So, these measures were based on the ACR 2012 guidelines. They were derived from that. That was based on a guideline review that did grade all the evidence.

DR. PACE: We understand that, and that is the reason that they can at least get

a moderate rating. But we have asked the developers, because all guidelines are different, we asked the developers to provide a summary of that systematic review of the evidence where they summarized the quantity, quality, and consistency from a systematic review.

And so, we understand from developers that a lot of times guideline developers don't make that available. But if it has been a guideline that is evidence-based, that is where we say, well, then, go ahead and at least give it a moderate rating. But, without really knowing the details of the systematic review, that is the thinking behind this algorithm.

MEMBER DANIELS: I feel like the young country fullback that is a freshman going against the UCLA Bruins All-American linebacker here.

But what I am going to say is I think moderate is okay. I mean, this

percentage of patients I think is something that has been shown.

And then, I will kind of make it complex and say, if they picked the six, I think that is probably a low. I will split moderate to low; you can pick which one you want.

CO-CHAIR CHOU: Yes, I think, again, we can address the targets when we get to the target measures. I think here essentially we are talking about do we think there is evidence support use of urate-lowering therapies in people with more severe gout. So, what we think the evidence behind that would be.

DR. PACE: And the question about where we are in the algorithm, actually, the answer to No. 3 is that there is evidence, but No. 4 is, is there a summary of the quantity, quality, and consistency? And that is where the "no" is. So, you are kind of looking in Box 6 now, based on the guideline

recommendation.

I also didn't see in this submission the definition of what those grades mean, which we can ask for. But it was stated it was grade A.

Each guideline developer tends to have different descriptions of what those mean, but the idea is, do you agree that that is sufficient evidence?

MEMBER BROTMAN: Can you provide an explanation of what "insufficient evidence," the conception, is and how it parses out at the end of that chart to a go-or-no-go status?

DR. PACE: Good question. I know that this gets a little complicated.

So, basically, what we think insufficient evidence means is, if it is based on more of an expert or consensus opinion versus actual empirical evidence, that is where we would say insufficient, versus low would mean there is evidence, but it indicates

you really shouldn't do whatever is being suggested. That is pretty unlikely that people will bring those kinds of things forward.

The other way you could get at a low is if the evidence is really inconsistent. So, there may be evidence -- and this is where the systematic review is very useful because there may be evidence, and, ultimately, when you look at the whole body of evidence, it may be inconsistent and conflicting. And the thinking there is, then, it is really not the right time to move forward with a performance measure.

But insufficient with exception would mean that it is primarily a consensus recommendation without a body of evidence behind it. But there are reasons that you, as the Steering Committee, thinks that it merits having a national performance measurement, which means, you know, maybe there isn't a better measure at this point in time; that the

recommendation is based on a group consensus; that that is the way to move forward versus just one small group, and the Steering Committee agrees.

So, someone brought up the question, well, why not just measure the uric acid? Maybe there is reason not to or that is too difficult. But you all are experts in the area and know where things are at and the possibilities.

MEMBER BROTMAN: Yes, just a followup. So, if that category is voted on, the same effect as low and insufficient evidence or --

DR. PACE: Yes. Okay, that is a good question. We have been kind of deciding how best to present this. We have done it different ways.

But low would mean it doesn't go forward. Insufficient evidence, it would not go forward. Insufficient with exception, it could go forward.

And if this becomes an issue, we may have to split it and first have you just vote on the evidence that is there, and then, come back and ask you to vote on whether you want to move forward with an exception. But we will kind of see how that goes.

CO-CHAIR CHOU: Yes, just to follow up on that, I mean, I have been involved with stuff with ACP, for example, with end-of-life care and counseling where there is no studies that show that that improves patient outcomes, but there are ethical and other reasons why people think that is a good thing to do. So, that is one possible example.

Another is in the area of safety, where we often don't have good studies. I have done a lot of work with opioids where we don't actually have any good data showing that doing urine drug screening and PDMP monitoring, and all this other stuff, actually reduces abuse or addiction rates. But people

think it is a good idea because people are dying from overdoses.

So, those are a couple of examples where we may not have good evidence, but enough of a consensus.

DR. PACE: And I think that is a good point. You know, the benefit should outweigh potential harm. And the reason we have this in here, and we should have mentioned this to begin with, is that these recommendations are for national standard performance measures, which puts in place, in motion, a whole infrastructure for people to make sure they are doing them, to collect the data, to report the data, et cetera.

And so, we want to make sure the evidence warrants, or in the cases that we talked about exception, warrants really putting that all into motion.

MEMBER GRAY: I would also like to ask a question, a clarifying question.

Roger just restated this: that

what we were voting on in terms of the evidence was that, if the medication was prescribed, it would lower the measures. But isn't the way this is written more focused on getting the doctors to prescribe it? Isn't that really kind of the key? Or which way does it provoke?

DR. FITZGERALD: So, if the patient has one of these three conditions and a high serum uric rate, then the doctor should prescribe the medication. And so, it is really about the doctor prescribing the medication under right conditions.

MEMBER GRAY: But do we have, I mean, in the evidence do we have that they don't prescribe it in those conditions?

DR. FITZGERALD: Yes, there is a big British study that is saying only 20 percent of patients might be getting that. It wasn't the exact same specifications. So, it wasn't a perfect fit. But there are lots of examples and citations where there is low use

of urate-lowering therapy.

DR. PACE: So, that would be about the performance gap.

CO-CHAIR CHOU: Yes. So, I was going to say that, I think, is the next step, opportunity for improvement.

For this first piece with the evidence, what we are looking for is that there is a link between an action and an outcome, right? So, we are looking for whether prescribing a therapy improves outcomes or reduces harms or both. So, I think that is where we are focusing on with the evidence.

We will talk about the performance gap I think next.

DR. PACE: I think you are asking, it is this what we talk about, the kind of pathway which has it has to be prescribed, but, then, as we have talked about, the patient has to take it before you are actually going to see the impact on the uric acid

levels.

Again, ideally, you want to measure things, the outcome or things closest to it, but this may be the best you can get at this point in time. So, you need to do that.

CO-CHAIR CHOU: So, I think we have one --

MEMBER MATUSZAK: I just want to make sure. So, this measurement that we are using here, we can only grade this as moderate if it meets grade criteria high or U.S. Preventive Services Task Force -- I'm sorry -- A levels, right? If it doesn't meet that criteria in either case, then we can't rate it as moderate? It has to go to low or insufficient?

DR. PACE: That is the way the algorithm is, that it really should be, without that quality, quantity, and consistency, that it should be based on that. But that is where the guidelines graded it Grade A. I don't know if you can speak to

what Grade A means for that guideline.

DR. FITZGERALD: We used the American College of Cardiology ratings for evidence A, B, or C. A was either a meta-analysis or more than one randomized controlled trial.

And so, in this instance, febuxostat studies would be two randomized controlled trials showing that the allopurinol data would probably be under that data as well. It is probably evidence B. It is multiple series.

DR. PACE: So, could you clarify, are you saying "you," as the measure developer, graded the evidence or the guidelines?

DR. FITZGERALD: The Guidelines Committee graded it.

DR. PACE: Okay. Oh, okay.

DR. FITZGERALD: Yes.

DR. PACE: Okay.

DR. FITZGERALD: So, the

guidelines would have been the recommendations related to the development of this being rated A based on the febuxostat studies.

CO-CHAIR CHOU: We have a moderate on the table.

One more comment here or?

MEMBER BROTMAN: I just wanted to state that, just for clarification, the rest of, I think 20 percent of the guidelines are based on level Grade A and the rest were Grade C, if I am not mistaken.

DR. FITZGERALD: I couldn't give you the percentages on the breakdown, but the majority of the gout data was B and C data. It certainly wasn't all C. But there was a lot of B, and the minority -- there are not a lot of gout randomized trials.

DR. YAZDANY: But I think it warrants repeating one more time that this reflects the fact that the data is old, and that the incentive to do a randomized controlled trial in the last 20 years on the

drug that costs cents is not there. And so, we just have to take that into consideration. It requires a nuanced look at the data. No one is going to do a randomized controlled trial of allopurinol at this point because there is no financial incentive to do so.

But that doesn't mean that the drug doesn't work. We have used it for 60 years, and it is just common knowledge and standard of care that it works.

CO-CHAIR CHOU: Yes, I mean, the strength of evidence, when we do our evidence reviews, we really look at that as a measure of certainty. And the way we think about it is, do we think if there are other trials that come out, will it change what our conclusions are? If somebody did a trial of allopurinol today, do we think it would tell us that this stuff doesn't work for gout?

And even if you don't have huge randomized, you know, placebo-controlled RCTs, you can still be fairly certain about that.

I have done reviews on vision screening for the Task Force where there is no trial of cataract surgery versus a placebo. You don't need one, right? You take out somebody's cataract and they can see, and they couldn't see before.

So, there's plenty of examples like that where, you know, we can use, I think, evidence that isn't as high on the hierarchy, or whatever, and still have pretty certainty of what the effects of the intervention is going to be.

I thought there was a question here.

MEMBER DODGE: I have concerns just about how the state of all the ledgers are and how contingent the evaluation of, say, this measure is on the others that actually didn't follow through on evaluating evidence and whether this actually helps people with gout as part of the short-term, especially if there is an increase in flares and people

initiate urate-lowering therapy. And if they aren't given these other prophylactic pieces, we could be doing net harm by just implementing this measure alone.

And so, how do we manage those contingencies when we are trying to take each measure as its own standalone?

DR. FITZGERALD: I think that is a good point. Again, I think in the more severe patients, the risks of not treating are a little bit higher. And again, we will work on educating. We will continue to try to do that, so that this is used in the most effective way.

I still think that in a patient who has had frequent attacks or evidence of damage, that getting them started on therapy -- those patients who are having attacks a lot, hopefully, they are on prophylaxis, and if they are not, there is a common custom on how to treat their attacks. So, I think in this select group, there is a

little less concern. I think in the group where there is little risk of progression to damage, then there would be harm. But that is not being applied to this measure.

CO-CHAIR TEMPLETON: This is Kim. If I could make a comment, too?

I guess making sure that the measure also is still focused on drug treatment when we know that the more holistic approach is to also look at lifestyle factors. So, would this potentially, then, result in less patient education about the other things that are part of gout care, with the full focus on medication/medicine?

DR. FITZGERALD: Lifestyle is important. There have been studies showing that lifestyle can lower uric acid levels as much a milligram per deciliter.

In fact, in the Japanese study that has been reported for patients with mild hyperuricemia, they recommend lifestyle.

The ACR guidelines do address

lifestyle and recommend that every patient should be counseled on lifestyle and changes.

For the subgroup of patients here with more progressive and advanced disease, there are studies that show that lifestyle alone is probably not efficacious enough in this subgroup.

CO-CHAIR CHOU: Yes, again, this would be another one where I would have a question about whether the denominator exclusion might incorporate whether lifestyle, you know, things that are identified to work first.

Again, I am not sure how strictly we have to adhere to the measure as presented.

DR. FITZGERALD: We really wanted to try to include -- that was both in the guidelines and in the measure development. But the complexity of trying to abstract from the chart lifestyle interventions and counseling we thought was going to be complicated. And so, we focused for a

simpler, cleaner measure.

CO-CHAIR CHOU: Yes. Okay. Well, again, we have a moderate on the table. Why don't we test the clickers? And then, we can see if there is a motion and make a vote on the evidence.

DR. PACE: Why don't you go back to just the one or two one? Yes. Yes, that one.

Okay. We are going to test the FOBs, so that we have got everybody's vote registered. And you have got 60 seconds to aim your FOBs.

CO-CHAIR TEMPLETON: And this is Kim. How do I vote?

MS. STREETER: Kim, if you want to text Chad and me through the webinar your vote?

CO-CHAIR TEMPLETON: Okay. Just hit the Send thing at the bottom?

MS. STREETER: Yes.

DR. PACE: And in this case, your

options are 1 or 2.

MS. PHILLIPS: Your options are 1 or 2.

CO-CHAIR TEMPLETON: And what is 1 again?

DR. BURSTIN: One is yes; two is --

DR. PACE: This is just a test. It really doesn't matter what it means.

CO-CHAIR CHOU: With your microphone on, repeat the instructions.

Just push 1 or 2 and point it to her.

MS. PHILLIPS: And point right here.

One more.

MEMBER GHOGAWALA: We don't have to say Send?

MS. PHILLIPS: No.

There we go; we have 22 now. Okay, so the voting works.

This is what a voting slide will

look like. So, back to the beginning of this.

MEMBER DANIELS: Can I ask, is it a strict majority? Is there a supermajority? What are the rules here?

Could I ask for a question before? Because I weighed-in. I wondered if the sophomore running backs wants to -- I'm pointing to Dr. Doge. Do you have an opinion on this?

MEMBER DODGE: My earlier comment was not just to the developer, but also to process, where if we are evaluating this measure and there are multiple contingencies about how likely this measure is to be effective that are based on other measures yet to be discussed, how can we evaluate that evidence?

MS. FRANKLIN: So, we still have to evaluate the measure just as it stands alone at this point and based on the input that we have received from the developer on their ability to modify or not modify the

measure.

CO-CHAIR CHOU: I guess, do we have a motion to vote on this as a moderate?

MS. FRANKLIN: Someone did ask whether this was a majority vote, and we do do voting by percentage. And if the vote falls within this 40-to-60-percent range, this is supposed to be reached and we consider the measure in the gray area. To the extent that it is 40 percent and above, the measure can continue through the process, to public comment, and then, consideration by the Steering Committee.

If the percentage of vote falls below 40 percent, then that would be for low or insufficient evidence. At this point the measure would not pass.

Are there still questions about the voting and the percentages?

(No response.)

Okay. I turn it over to Dr. Chou.

CO-CHAIR CHOU: All right. Do we

have a motion to vote on the evidence?

We don't need to? Okay, let's just do it.

MS. FRANKLIN: So, for this measure, you would be voting high, moderate, low, insufficient. And then, if the Committee feels like we should exercise an exception to the evidence, if there is an insufficient vote, then we would look at --

DR. PACE: Well, no. Right now, let's try this, unless we get caught up on the insufficients.

But, basically, 1 is high, and we have already kind of talked about that, only eligible if the summary of the quantity, quality, consistency submitted -- 2 is moderate; 3 is low, meaning the evidence would really indicate you shouldn't do what is being suggested. Four would be insufficient evidence, but you think it meets the exception, that even though the evidence isn't strong, it really is something that, you know,

for consensus opinion and benefits outweigh harms. And then, 5 would just be there's insufficient to make a decision one way or the other. You just can't tell because there wasn't enough evidence. Okay?

MS. PHILLIPS: We will have 60 seconds, and the voting starts now.

(Vote.)

MS. PHILLIPS: We will be taking 22. We have one remote who is voting via chat. That's right, we are 21; that is correct.

So, we can see that 67 percent is moderate. Low was 10 percent. Okay.

Fourteen, moderate; 2, low; insufficient evidence with exception, 1, and insufficient evidence, 4. And that means the measure can go on.

CO-CHAIR CHOU: So, I think we now move into this performance gap issue, the opportunity for improvement.

Dr. Daniels or Dr. Dodge, would

you like to present? We have already talked about it a little bit, but maybe just summarize it real quick?

MEMBER DANIELS: Well, basically, you know, the new evidence that they have here, there is probably some room to improve that, you know, with patient compliance.

It is pretty clear. So, I guess I will keep it short. That's all I will say.

Basically, the Work Group came out and said that, overall, less than optimal performance of quality of care; providers will kind of already know. And there is disparity among groups, and then, with the information from the study and, also, the UK information --

CO-CHAIR CHOU: I guess my only question is that most of the data presented seem to be about adherence and not about prescribing rates, but I assume that they are both an issue.

DR. FITZGERALD: Yes, I would

think the evidence that best addresses the prescribing rates is what I would provided in the handout by Quo and colleagues. It was a 2013 publication, and it was looking at the trends of the incidence and prevalence of gout, which is the two left charts.

And the chart in the right column shows the proportion of patients that had been prescribed urate-lowering therapy. That has remained low. These are patients who are diagnosed with gout. That has remained low, and it has actually even fallen off over recent years.

CO-CHAIR CHOU: Any other discussion or questions or comments?

(No response.)

All right. It sounds to me like it is time to vote on the performance gap issue here.

MS. PHILLIPS: You have four options, 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

You may begin voting now.

(Vote.)

We have 20. We need one more. There we go.

Okay. We have 4 for high. We have 15 for moderate. We have zero for low, and we have 2 for insufficient.

CO-CHAIR CHOU: So, I think that means that we agree that this is high-enough priority in terms -- or excuse me -- there's enough opportunity for improvement to move on. Correct?

All right. So now, we are into the priority discussion. And again, Dr. Daniels or Dr. Dodge, could you briefly tell us what you think about how this measure addresses a healthcare priority? Again, we have heard a lot of this before. So, it can be pretty quick.

MEMBER DANIELS: It's there. You know, the stuff that I read from, actually, stuff they gave me last night, some of the

things that I thought were linked aren't, or at least there's not good evidence. You know, I always thought that the metabolic syndrome and all that really affected it, and there really isn't any direct evidence. There is some association, but there is no direct evidence. But there is pretty good evidence that there are problems with the tophi and that.

The number of patients, you know, how big this hits on the National Priority, I kind of bow to the group here. It probably knows more than I on what you need to do, on how many people this will affect when they come into the doctor's office. By the time they go through the things they want to go through, you want to go through, and then, what the guidelines go through, is there enough time?

So, I will leave it at that.

MEMBER DODGE: I think the only addition -- and this came up on the call -- is

that there was a study or there were numbers about gout in general, but not necessarily what subgroup this particular criteria represents of that 8 million Americans that suffer from gout. I think 20 to 30 percent sounds like the estimate, which is still is substantial. But that wasn't clearly specified.

CO-CHAIR CHOU: Any other questions or comments?

(No response.)

All right. I think we are ready for a vote here. So, this is on healthcare priority.

MS. PHILLIPS: You have four options, 1 for high, 2 for moderate, 3 for low, and 4, insufficient.

And the voting begins now.

(Vote.)

And we are at 21.

We have 1 for high, 14 for moderate, 3 for low -- I'm sorry -- 2 for low,

and 4 for insufficient.

CO-CHAIR CHOU: So, I think it passes for that criterion as well.

So, then, I think we move on to kind of more the kind of implementation stuff. I don't think the quality construct is relevant for this measure.

And so, the next area that we need to address is the reliability. The testing isn't relevant, right, because it hasn't been --

DR. PACE: Right, and good point; we probably should have added a new slide for the trial measure.

So, basically, what we need to look at here are it is HQMF specifications.

And did Chris or the team look at the specs and say that they were sufficient? I am not sure we need to go beyond that, but, Angela, I am not sure what has been done.

MS. FRANKLIN: I believe, based on our review of the specifications, the HQMF

specifications, they were sufficient with the measure.

DR. PACE: And that's confirmed?

MS. FRANKLIN: Yes. So, that was what we look at.

CO-CHAIR CHOU: So, does the Committee vote? I mean, I am not clear exactly what we are voting on.

DR. PACE: So, it is a good question. So, it is an excellent question. We need to think this through here and ask you to think it through with us.

So, I guess what we could do, you know, we did have staff that reviewed the HQMF specifications. We don't really expect all the Steering Committee members to be able to do that.

We will be asking you to vote on feasibility, which gets into the feasibility of the data elements of the measure logic, which the developer should have presented something for you.

So, I guess what we could do is just ask if anyone had any questions about the HQMF specifications or any concerns about the specifications that we should address. But I am not sure -- I will ask you if anyone has anyone has any suggestions. I am not sure that we would really need to vote on those. But were there any issues with them?

Yes?

MEMBER VISCO: I guess maybe just a point of clarity then. Since there is nothing under exclusion criteria ends, and I know there is some wording in there regarding the specific -- if a patient declines a medication, then there has to be some evidence behind that.

And again, I am not as familiar with this evidence. And if there is evidence behind patients deferring or declining medication, you know, they want to put arnic on their feet for three weeks before you give them the prescription, or whatever.

And then, the second thing would be contraindications to oral administration of medication.

DR. PACE: So, I think you raise some good points. I think we should have the Committee at least look at the English language of the specification and raise any issues you have about it, exclusions and things of that nature. So that we can make sure that those are clarified and we can think about voting regarding this while you are talking about them.

But, you know, there really are two issues about specifications. One is that they are precise, and that is part of what the HQMF review would do, and, also, that they are in appropriate HQMF format.

But the other part is, are the specifications consistent with the evidence that was presented? And I think that is part of what you are getting at.

So, why don't you just have a

brief conversations about those issues?

CO-CHAIR CHOU: So, it looked to me like the reliability is just whether they report what they want us to measure. But the validity is where we get into this evidence, you know, what the match is and whether there should be different exclusions, et cetera.

DR. PACE: Right. So, I think what we are just going to do is focus on specifications. And there are two issues about specifications. It is really do you have any questions about the specifications that need to be clarified. We already know that the HQMF works, but if you have questions about the specifications and whether they are appropriate, then I think we should bring them up here.

CO-CHAIR CHOU: Right. So, the one thing that came up earlier was that people with frequent attacks, it will only be people who basically present to the office and the doctor codes it as a gout attack. So, there

will be probably a lot of people that either don't come in or it is not coded, and they are going to be missed in the denominator, correct?

DR. FITZGERALD: Correct.

CO-CHAIR CHOU: Okay. And then, we had already talked earlier about just the question about how somebody defines a gout attack or clinician, but I don't think we are really going to be able to get into that here.

Are there other questions or issues in terms of the specifications?

MEMBER ANNASWAMY: Validity of including -- there is no denominator exclusion specified in this measure. So, as my colleague here mentioned, shouldn't we exclude patients who voluntarily say, "No, I don't want to be treated with drugs."? Or are there contraindications to them? Are those denominator exclusions?

DR. YAZDANY: So, traditionally -- and Karen can correct me if I am wrong -- but

patient preference, we have guidance that that is not an appropriate exclusion, based on NQF materials and sort of the national climate.

DR. PACE: So, exclusions or exceptions has gone back and forth. And so, the NQF criterion guidance on the patient preference really shouldn't be part of exclusions unless they can be appropriately identified, because that is something -- a lot of this comes from the idea, you know, it is something that is easy to check off.

You know, the question with exclusions is really, is there any reason to think it should be different across providers? So, things that are infrequent, kind of random events, or that there's going to be a certain small percentage across providers, where that happens, then it is not going to appreciably affect the measure.

MEMBER ANNASWAMY: Can you give an example?

DR. PACE: Of a small random thing

or --

MEMBER ANNASWAMY: Of an exclusion that is not a patient preference?

DR. PACE: Yes. When you are talking about drugs and allergy to drugs. And that is directly supported by the evidence, I mean the clinical evidence.

And again, I guess one could argue that that is going to be a small percentage. And that is part of what, you know, when they actually bring a measure forward for endorsement, if there are exclusions, we ask for some analysis of those exclusions.

But my understanding with eMeasure is also that these kind of broad, general categories are very difficult to specify as eMeasures. So, in the past we have seen with some of the measures like a general category of patient preference, system issues, or medical reasons, without any specificity.

First of all, you have a problem with standardization, but, also, they have

been very hard to operationalize in an eMeasure environment.

But go ahead.

MEMBER DANIELS: I just want to chime-in that I agree with you, Roger, that really the crux of this whole thing is in this validity area that I hope gets sort of sorted out, because I think there are lots of studies, but they don't actually hone-in on the exact question that we are doing. So, that would be the part, at least for me, that would be real important.

CO-CHAIR CHOU: Right.

I was just going to comment that, you know, ACR I don't believe has adopted grade yet, but the way that grade rates the recommendations is that strong recommendations are really not affected by patient preferences. I mean, they are, basically, we think this is what should be done pretty much all the time, unless there are really extenuating circumstances.

Whereas, the weak recommendations are much more conditional or based on shared decisionmaking kind of aspects. And those are ones where grades would actually say they are probably not candidates for quality measures. Because if patient preferences are that big of an issue, then we probably shouldn't be measuring clinicians on how they perform on this.

So, in general, we want the quality measures to be the equivalent of a strong recommendation. So, if there is a concern that there is a huge patient preference component, I think it does call into question whether it should be a quality improvement measure.

MEMBER ANNASWAMY: How about contraindications?

DR. PACE: Yes, I think those are certainly fair game for exclusions.

But do you want to address that?

DR. FITZGERALD: So, with urate-

lowering therapy, there are three options: the allopurinol, febuxostat, and the probenecid. And each of them has some specific inclusions, but we had in our prophylaxis measure, because the prevalence of those contraindications are higher, we listed all those.

For this one, allopurinol hypersensitivity is really fairly low. Febuxostat is an option for those patients. Probenecid has contraindications. Probenecid is not used all that frequently. So, we didn't specify those. We can certainly look into those areas during the testing period.

MEMBER ANNASWAMY: Even if there are contraindications? Is that something you can capture by an eMeasure?

DR. FITZGERALD: We will look into that, but we should be able to capture, for example, an allergy for allopurinol. For probenecid, we could capture a creatinine level or a history of stones. Febuxostat,

again, an allergy would be something, but the allergy for febuxostat and allopurinol is quite low, the allopurinol clearly more concern. The febuxostat can have some similar outcomes.

CO-CHAIR CHOU: So, again, a process question. So, if we do think that there should be a contraindication exclusion or something, that is okay?

MS. FRANKLIN: Yes. We would capture that in our discussion that will be sent out for public review and our report.

So, the concerns that are being raised around the table would be included in our evaluation of this measure, although we are not voting on this particular criterion at this time, and be open for public comment.

CO-CHAIR CHOU: Okay. So, I think we are really in the validity discussion. I mean, I think we have kind of moved a little bit beyond the reliability.

And again, I think the discussion

is around the denominator and the exclusions and these kinds of things, whether we think it is specified the way that we wanted to or whether we have questions or comments to make.

So, I will open up the floor and see if people have questions or comments here.

DR. PACE: Right. Very good.

So, this is a measure that is coming to us that we are considering for this approval as a trial measure, which means it really has not had formal reliability and validity testing.

So, the only thing that you would be voting on are specifications. So, I guess what we could do -- would you move to the reliability slide.

So, in this case, maybe we can do it this way: you would only be voting on 2a1, precise specifications. So, that is part of our reliability criteria. And normally, reliability is mostly hinged on what the testing actually shows. In this case, there

is no testing.

We have had a review of the HQMF specs, and those are sufficient. So, I would say that --

MS. FRANKLIN: The voting speaks only to the human-readable specifications --

DR. PACE: Right.

MS. FRANKLIN: -- we have in the measure as specified now. Does that make sense?

DR. PACE: Yes. Or we could just skip these and just offer comments. I think that is probably a more reasonable thing to do.

MS. FRANKLIN: Yes.

DR. PACE: So, let's just say let's move on and talk about the specifications in terms of whether you think there are any issues with them being the right specifications, which normally would fall under validity.

But we are not going to really

have you vote on either reliability or validity, but certainly offer any suggestions, because these measures are going to be approved as trial measures, which means they are going to be tested and, then, eventually brought back for full endorsement, where you would be actually looking at the empirical reliability and validity testing.

So, does that make sense to people? And I apologize, we are kind of working through this with you as we are looking at this first measure for a trial measure. Okay.

CO-CHAIR CHOU: Comments or questions?

(No response.)

So, I still have some concerns about the potential exclusions in the contraindications piece. So, I would suggest that that is something that should be added, or at least I would like for that to be added.

And then, I also wonder if there

should be built in some kind of lifestyle trial or something. I mean, this basically obligates clinicians to start treatment, the drug. And I wonder if there should be some period at least where you are warned in lifestyle and other therapies. So, that would be my other concern.

Then, my last thing that I brought up before is the denominator. I mean, I am fairly comfortable with what it is going to be for tophi and erosions. I am not quite as comfortable with drug therapy for everybody that has a couple of minor attacks a year, however they are defined.

And so, I think I can live with it because of the terms of what the denominator should be. I just wanted to put that on the table.

I think you had a comment?

MEMBER VENTURA: I was under the impression that the denominator captures that by only addressing the people that are more

severe, because you have to have the serum urate 6 or greater plus the two frequent attacks.

CO-CHAIR CHOU: I don't think it specifies the uric acid level.

DR. FITZGERALD: It does.

CO-CHAIR CHOU: It does? Okay. So, a serum uric acid of 6 or greater. The tophi and erosions are specified.

So, I guess if people are comfortable with a uric acid level of 6 with people who have monoarticular arthritis a couple of times a year or what we think is monoarticular arthritis, that is fine. Like I said, I just wanted to put it on the table.

MEMBER DANIELS: I would like you to split it because I think we have sort of two issues here. If you put the gouty erosion and the tophi, that is sort of looking at like long-term damage type of thing versus the gouty flares. But even if you look at treatment, you know, in practice, we have got

people who have gouty flares often getting anti-flammatory medication. It is sort of like a different pathway.

So, I would almost like to see it clear to me. And then, I would have a lot more comfort with the levels you have taken because, you know, I do actually have -- people who have the gouty flares, that was recommended for a whole series. The reason that I had read is that the complication rate was higher with the higher levels. So, the higher levels, like maybe up around 8 or 9, may be associated with the things that could really affect people's health, not that having a gouty attack isn't fun. I understand.

To me, that would really clean it up as far as being able to sort of support it with the evidence. Because, right now, it is just like a whole bunch of targets kind of mixed up in there. And it is not your fault, but the studies that you have don't exactly address those sort of perfectly. And so, you

are making some inferences. So, the cleaner we can get it, the better I would feel.

DR. FITZGERALD: So, you're right, the studies that identify patients with poor outcomes, tophus is a poor outcome; high serum urate is a predictor of poor outcome and higher medical costs.

Patients who are having frequent attacks are more likely to have serum urates in those high levels. I think we are thinking about the odd character who is having frequent attacks and has a marginal urate. Those people are just less common.

If you are having to address the recurrent podagra, if you are having frequent attacks, the natural history of gout, it progresses from an intermittent basis of attacks to a chronic phase, where there starts to become chronic pain, chronic inflammation, damage, and erosion.

And the people who are more likely to go on to that are those who are having the

frequent attacks.

MEMBER DANIELS: And again, I know the waters will get muddy a lot more in the future because, as your specialty has sort of done, now the residents are getting trained on office ultrasound on a lot of people. And so, that is going to change like the sensitivity of that. Right now, it is kind of hard to pick that up. So, I think that in the future that may become an issue.

DR. FITZGERALD: It may. Again, gout is still primarily treated in the primary care office. Eighty percent of patients don't make it to a rheumatologist.

MEMBER ANNASWAMY: There was a mention of erosions in the inclusion criteria, but it is not actually on your enumerator statement. And you just mentioned tophi in your enumerator statement. Is there a difference between tophi and erosions?

DR. FITZGERALD: In the last revision I guess erosions came out.

MS. MYSLINSKI: So, when we were working on the e-specification, it was not really feasible to assess that in an electronic way, those erosions on a radiograph.

DR. PACE: So, I think the question that you are asking is whether the specifications are consistent with the evidence. So, for example, does the guideline for giving this med or prescribing med, is it specified for these specifications?

DR. FITZGERALD: The guideline from the 2012 ACR guidelines is -- and, actually, so have been the other guidelines, including the British Society and EULAR -- has been two or more attacks per year tophi or erosions.

MEMBER GRAY: Just one more point of question here. Is the target, then, for the accountability to the physician, it is really primary care then as opposed to -- like once they get to the rheumatologist, you

believe they are prescribing, right? Is that sort of the deal, the 20 percent that is going there?

DR. FITZGERALD: I think assumptions about rheumatologists doing things properly are assumptions. And there is some data that there is better adherence and more prescription use, but you would expect that. They are also a different group of patients that is being seen. So, it is a little bit of apples and oranges.

But the goal of these measures is really to try to improve the care for all gout patients. And since most gout patients are being seen by the primary care doctor, that is the target.

Again, the infrequent minor gout patient is not being targeted by the measure, but, hopefully, their quality of care improves as well.

MEMBER ANNASWAMY: A couple of questions. One is age 18, is that specified

because less than 18 doesn't happen? That is more of a clarification.

And the second is, what does the lead discussant have in terms of the motion for the specification?

DR. FITZGERALD: In general, under age 18 is very unusual, yes.

MEMBER DANIELS: And I guess this is where the country boy gets hammered by the outside linebacker.

I am going to put 3 down for this. This is where I have my issues with it.

MS. FRANKLIN: Okay. So, at this time, we are not actually voting, but we can capture the Committee's sense right now in the discussion about where -- we are not looking at feasibility; we are looking at reliability and validity -- validity at this point.

DR. PACE: The question is, though, ultimately, you are going to be asked to approve this as a trial measure. So, if you think the measure is not specified

correctly, then I guess that would be a stopping point.

So, do you agree that we need to have them vote on this?

CO-CHAIR CHOU: Yes, maybe let's take a second. Do people -- I mean, I have already said some of my concerns, and I think James has also -- do other people have thoughts about whether some of the specifications, we should consider modifications, or want to support the specifications as is?

MEMBER GRAY: Can we hear from the Committee or the key people, JD and Christian?

CO-CHAIR CHOU: Okay. Who? Chris?

MEMBER DANIELS: Yes. Basically, I would feel much more comfortable -- I will just repeat it -- if they sort of split it, if like this gouty flare was sort of thrown out and they are looking at erosions and tophi, because that is kind of more of a problem.

And the other issue I have is with the level. I almost wish we wouldn't check in that particular level because I think that opens up a whole can of worms.

You know, those are my two issues, and the other thing is that they have got some big studies here, but they are not exactly asking the same question. They are close. So, there is a lot of sort of inference in this. That is why I am putting it low.

CO-CHAIR CHOU: Thank you.

John?

DR. FITZGERALD: So, we can feed back after we get our specification. When we test these measures, we could feed back the differences between gouty flare being included or excluded. We could look at that.

And then, the reason to put a 6-milligram threshold is we don't want to ding -- so, people who have tophi and gouty attacks, even if their uric acid is 4.8, they should still be put on therapy. But that,

again, is sort of that "rare hen" that we don't want to ding people for. So, we put a minimum threshold to make sure we weren't expecting people to go beyond that.

MEMBER DANIELS: Then, on 2549, you are using 6.8, on your other. That is where the confusion is.

DR. FITZGERALD: Yes, I can go into -- there was a lot of data on that -- and I can go into the rationale on it.

On that one, we didn't want to hold people to try to get -- one is an indication to use the prescription; the other one is a target. And so, a higher target is more lenient. And so, we chose a higher target with rationale behind that, to be more tolerant of the patient who is having two attacks per year, and they come back with a uric acid of 6.2. We didn't want to ding someone for that.

CO-CHAIR CHOU: Again, we are going to talk about the targets later. I

don't want to get too caught up in that.

But there is one thing I noticed when I looked at the Becker trial, the enrollment criteria was you had to have a uric acid of over 8. And so, again, this uric acid level thing is just all over the place.

DR. FITZGERALD: Okay. So, just my comment on the outcome to try to get the drug approved was serum urate. So, the pharmaceutical company is going to want to enroll people who have high serum urates. They are not going to enroll people with, you know, a serum urate of 6.8 with a goal of trying to get an answer. This affects their image and their population.

CO-CHAIR CHOU: No, I understand that. I was just trying to connect the evidence with what we are actually recommending in the quality measure.

DR. FITZGERALD: Again, trying to get the evidence to exactly match this, the wording and the specificity and the levels,

there is going to be a limit on what the randomized trials can provide.

CO-CHAIR CHOU: Yes?

MEMBER BRYAN: I'm sorry, getting back to JD's comment earlier about the gouty erosions, it seems like in the specifications it is strictly related to radiographic erosions, but it can also seen on ultrasound and MRI.

DR. FITZGERALD: So, Rachel reminds me that we put the erosions at the very end, when we started looking at the feasibility.

To answer that, the guidelines have looked at erosion, ultrasound and dual-energy CT findings, and those are really too new to put in and expect people to use as targets. And so, those were left out of the measure.

CO-CHAIR CHOU: All right. So, I guess I need to get some guidance on what we are doing now.

(Laughter.)

So, we are now voting, I heard.

DR. PACE: So, again, you will be ultimately voting at the end on whether this measure should go forward as a trial measure, meaning that it is ready to do testing.

So, I think there was a suggestion that the denominator specification should be looked at, and the developer mentioned that that is something that could be analyzed through their data when they obtain that through testing.

So, one way we can do this is for you to offer recommendations in terms of what the testing should or some analysis that you might want to take a look at with the specifications.

I think what we might do is move forward through the other criteria. At the end, you will be asked to do a yes/no on approval as a trial measure, meaning it still has to go to testing and would have to come

back to NQF with testing results in order for it to pass endorsement.

So, that is where you will really weigh-in on reliability and validity, but I think this is a good opportunity to tell the developer the kinds of issues or concerns you have, so that they adequately look at it during their testing period.

Does that satisfy people, if we move on to the other criterion?

Go ahead.

CO-CHAIR CHOU: Yes, I mean, I think at least James and I have expressed some suggestions about analyzing the people with recurrent attacks separately, however you guys want to do. I think getting data separately would be good. The contraindications for it in terms of exclusions would be something else to consider.

And then, at least in the people without tophaceous gout and erosions, considering whether you can somehow

incorporate this kind of non-drug therapy trial, or whatever, would be my other suggestion.

Are there others?

MEMBER DANIELS: I'm sorry. And just to clarify, maybe even considering just saying we are just testing, not asking seeing what -- just kind of tying two things in there in a way, just to see if they are measuring it. It is not actually saying it has to be this level.

CO-CHAIR CHOU: Thank you.

All right. Are we supposed to break or do we want to finish this one?

DR. PACE: Let's go through this, because I think we should be able to get, and we need to get, through these last.

Let's move on to feasibility.

CO-CHAIR CHOU: Okay.

DR. PACE: And then, we will do usability and use, and wrap this up and take a break.

Is that okay, Angela?

MS. FRANKLIN: Yes.

DR. PACE: Okay.

CO-CHAIR CHOU: James, any comments about feasibility? I mean, personally, it is hard, it is, I think, more difficult for the panel members who will judge some of these, because these are more kind of technical issues. I think we would like to hear if you guys think there are feasibility issues as well.

DR. PACE: So, Angela, did Chris review the feasibility?

MS. FRANKLIN: I believe we found the feasibility to be sufficient, and the developer also provided additional information around feasibility assessment that they can docket for the Committee to review.

And I guess we would ask the discussants if they have any discussion that they have or comments that they have regarding feasibility, and then, concerns or questions

from the Committee.

MEMBER GRAY: I have one, and that has to do with, since the ICD-10 will be put off for another year, how does that impact your ability to clarify this data?

DR. FITZGERALD: So, we have picked testing sites based on ICD-10 implementations. And they were all ahead of the deadlines. And I don't know if we have had updates since the new factors.

CO-CHAIR CHOU: Yes?

MEMBER VENTURA: Can I ask the developer about a feasibility table, an Excel spreadsheet? I am not sure what those numbers mean.

DR. FITZGERALD: So, a survey was done of the sites, and there were questions to each of the leaders at the various test sites about the ability to abstract the data from their records.

MS. FRANCISCO: So, I think a primary issue is that whatever you were

looking at during the call that we were on was not the actual feasibility assessment that we submitted prior to the discussion. And we actually submitted an entire --

CO-CHAIR CHOU: Can you speak closer to the microphone?

MS. FRANCISCO: I'm sorry.

We actually submitted an entire workbook that runs through each of the questions that were asked, each of the data elements that we asked the sites to look at, and a summary of feasibility based on all those questions that we asked on the specific data elements.

So, we sent that in. We have resubmitted it, and I am hoping that you have all actually have had a chance to look at that.

MEMBER BROTMAN: So, does the scoring table, the results that are in that -- it was a 3x3 grid -- does that have an explanation of what those numbers mean?

MS. FRANCISCO: You will find the guidance provided by NQF in terms of the feasibility numbers. We sent the feasibility assessment. And I am not sure if I can quote them off the top of my head. But I think we included in the worksheet the definitions, maybe in the overview tab.

DR. FITZGERALD: Yes, Dr. Ventura just has that little final table, not the explanations.

MEMBER BROTMAN: We only have numbers, 1, 3, 2.

MS. FRANCISCO: The very small table that you are referring to, I think it is a 3x3 that has an overview summary. That was submitted very, very early on, before we went through this entire process. So, this document was uploaded and submitted to you for review, and I don't know why, but you were all given the original table that we submitted very, very early on. So, these are the actual results from the full feasibility assessment

that we conducted.

DR. YAZDANY: Melissa, would you just provide an overview of the results, because it sounds like maybe people have not had a chance to look at it. So, just a high-level summary, maybe the high points?

MS. FRANCISCO: Sure. I think it would be helpful if you could just put the overview tab up. Then, I can kind of speak to that.

So, we went through testing of critical data elements that were identified, that we needed to assess the feasibility of, in order for us to even proceed with testing of these measures as they are specified.

And so, you don't all have to look at the screen. We looked at five critical data elements relating to certain comorbid conditions, chronic kidney disease and diabetes. We looked for a diagnosis/finding for tophus and gout attacks, contraindications for a certain list of medications that are up

there, as well as lab results for serum urate, and active medications, ULT and NSAIDs, the corticoids and colchicine, with a standard set of questions that we asked that are based on NQF.

And the rating scale that we have included on each of these data elements also is based on NQF's standards and speaks to feasibility rated 1 through 3, based on, "Is this data element feasible to collect at this time? And if not, is it feasible to collect," I believe it is within the next three to five years.

CO-CHAIR CHOU: Okay. So, it seems like we are getting there, I mean in terms of the general directions to go. And so, hopefully, it will be helpful for our gap discussion, too. Measures like the Minnesota total cost of care endorsed measure do get at those buckets that you were just describing.

In addition, we have talked about breaking those measures down by payments from

third-party sources.

And in addition, we have also talked about the episode-based measures, the measures that are more centered on location, doc conditions, and specific applications of care that make up these buckets.

So, this is helpful for an overview framing, and we do have, can look at least the top two measures, the total-cost-of-care measure and the knee replacement measure that both fit into this framework. So, this is getting somewhere.

MS. FRANCISCO: So, there are specific ICD-10 codes. If you look through the value sets, there is an extensive list of codes just on tophus alone that get into the granular detail of how that is defined.

CO-CHAIR CHOU: And we expect people will be using to that degree of specificity? Anyway, I guess we don't get into all of that. But that is one of the questions I have. Are we actually really

going to be able to identify people with erosions and tophi?

MEMBER MATUSZAK: I would just also submit that I don't know if we can determine feasibility if we still have a lot of questions about what the measure is actually going to include and what it is not going to include.

If we start talking about more exclusions and they are rewriting parts of it to maybe handle more exclusions, then do we really have a good handle on how feasible it might be, once you have started to add some more into the denominator?

DR. FITZGERALD: We will certainly test for the exclusions, as requested. We are going to be collecting the data anyway.

Some of the exclusions, you know, the contraindications, those are going to be fairly infrequent. It is an infrequent event over multiple providers. I don't think it is going to be a concern.

So, I don't want to speak for NQF, but I would vote on it as --

DR. PACE: Yes, I would vote on it as it is specified. You know, the developers is going to take your suggestions to heart when they do their testing. This measure is not going to be endorsed. You're basically saying this measure is ready to go and be tested, and you are offering some suggestions for additional information that you need to, that you would expect to see when this comes back.

This is a standing Committee, which means that this is the Committee it will come back to.

So, I think when the testing comes back, if it indicates they hardly identified any patients, that will tell you something. Perhaps it won't even come back if it really doesn't make it sufficiently through their testing.

MEMBER JARRETT: Yes, it gets to

the issue, you know, if I look at the table, a lot of the feasibility is three to five years, which is a long time in the electronic world that we are living in today because everything will be different.

And philosophically, I mean, what would be the timeframe that the testing would occur? Because, clearly, if it is three to five years before they will be able to pull down a lot of this data from large groups of populations, then are we setting something up now that may just completely change in three years again because of the fact that the electronic world is changing?

DR. FITZGERALD: We do have a calendar. Do you want to describe that, Rachel, the testing? Maybe go to the history with like RA and other experiences.

MS. MYSLINSKI: So, we are anticipating the testing over the next 12 months, so over the next year, which is consistent with RA testing.

CO-CHAIR CHOU: Sorry, I was going to say, the one other thing, just in terms of feasibility, is I am just not sure how well you will be able to distinguish what an acute gout attack is. Because, again, when I see patients with gout, I have just coded it "gout". And it could be just a followup or it could be an acute attack. I don't think you can tell necessarily from the coding. So, just another thing.

DR. FITZGERALD: I think with ICD-9 you are limited on that, but ICD-10 does have specific gout, acute gout, as a code. So, if you were giving someone an NSAID prescription for acute gout management, in the future you will be coding that with ICD-10 for gout.

And the sites we are going to be looking at, it is going to be primary care practices who are using ICD-10.

CO-CHAIR CHOU: So, I guess we are voting -- oh, sorry.

MEMBER ANNASWAMY: Well, that brings to question about whether you're voting on whether it is feasible now or feasible when ICD-10 comes in.

DR. PACE: It is specified for ICD-10. So, it is an eMeasure specified for ICD-10, and that is what the feasibility assessment reflects. And they are going to test it in sites that are using ICD-10.

MEMBER ANNASWAMY: To borrow the football analogy, we are voting on potential, not on --

DR. PACE: Yes, yes, exactly, because this is an untested measure. So, basically, the bottom line for these trial measures is that it meets our importance criterion, because there is no point in going forward if it doesn't meet our importance criterion, even for testing.

And they have already done their homework on feasibility, that they think that they can actually implement it in these test

sites, so that they can accumulate data to do formal testing on reliability and validity.

And that under usability and use, when we get to it, it is that, basically, there would be a use for this in improvement and accountability, and what is the plan for that to put it into use.

And all of this endorsement will, then, hinge on whether the testing actually demonstrates that it can be a reliable and valid quality indicator.

CO-CHAIR CHOU: Any other questions or comments?

(No response.)

So, I guess, if there are none, we will be voting on whether we think it is feasible enough to test.

So, I guess we will go ahead with the vote.

MS. PHILLIPS: Okay. We are voting on feasibility. You have got four options. One is high; 2 is moderate; 3 is

low, and 4 is insufficient.

You may begin voting now.

(Vote.)

MS. PHILLIPS: We have 14 for moderate. We have 5 for low, and we have 2 for insufficient.

CO-CHAIR CHOU: I believe that passes.

And our last issue or criterion is this usability and use.

Some of these I think aren't really relevant in terms of we know the measure isn't being currently used and it is not being publicly reported. And we don't know if there is going to be any improvement over time since it hasn't been implemented yet.

I guess the one major issue would be this unintended consequences piece, which we have talked about before.

Would the lead discussants want to comment on those before we open it up to the

rest of the panel?

MEMBER DANIELS: No.

DR. PACE: The other thing is whether there is a rationale how it could be used for improvement and that there is a plan for it to be used in accountability.

DR. YAZDANY: So, I can comment on that.

The ACR has been collaborating with the people who developed the American College of Cardiology's PINNACLE Ambulatory Care Registry. So, I think that is one of the successful examples of an ambulatory registry in the United States.

They are in the process of scaling our Pennsylvania Registry, which is gotten certification by CMS to be a Qualified Clinical Data Registry. And these gout measures will be part of beta data, eventually a part of that program in the coming year, we are hoping, and the measures have also been submitted to CMS for use in 2016 programs.

So, on many different levels, we are trying to get more experience with the measures and trying to get more data on gaps in care and potential for improvement.

CO-CHAIR CHOU: Thanks.

Yes?

MEMBER MATUSZAK: I want to echo what you brought up earlier, Roger. But I do think that there is a high potential for unintended negative consequences with this. You are talking about that you are specifically gearing this towards the practice in primary care providers and the community.

And now, you are telling them that we are going to judge your quality based on whether or not you are starting urate-lowering therapy on the cost of at least two gouty attacks, and we are going to grade you on this.

I understand right now we are just testing, actually, but now you are probably going to have a significant in the number of

people who are experiencing untoward effects from the initial period of time, that they are not getting adequately educated and just being thrown en masse onto these meds to try to meet these numbers for the quality measure. I think we will see a significant increase in the number of people a lot of side effects or a lot of increased flares during that acute period of time because of inadequate education.

CO-CHAIR CHOU: Yes, I mean, I think I said something similar before, and I think the other aspect that I am concerned with is unnecessary treatment and this issue of really pushing people to drugs without considering other lifestyle stuff first.

I know that NQF, they must have dealt with this with things like diabetes management and hypertension and hyperlipidemia, because these are all conditions where you would like to do lifestyle stuff first.

So, there must be precedent for this kind of thing. But I do think it is worth considering.

I think one of the challenges will be how do you actually measure the unintended consequences. I am not sure how we can do that, which is a little bit of a concern because we would like to be able to know what the consequences are. But I think it is actually going to be very hard to measure that.

DR. BURSTIN: Just one more caution. The criterion actually makes the case that there is evidence of unintended consequences. So, we are really talking theoretical.

I think your point is really well-taken, though, Roger. We need to have a better system of understanding feedback on the ground. Perhaps that is something that is getting better, because our registries are really beginning to understand that, and

perhaps put in what some people call "balancing measures" to kind of keep an eye on what might be measures that would suggest that there might be a problem beginning, as part of your ongoing surveillance.

CO-CHAIR CHOU: Yes. Well, there's a couple of people I wanted to comment.

I was going to say there could actually be a measure about, is lifestyle stuff tried first? I mean, that could actually be a separate measure, but we don't need to talk about that right now.

So, a couple of comments. I think you had one.

MEMBER BRYAN: It is just to dovetail on what Jason said and what you had said earlier, Roger, the concern that how many of these folks are being labeled with gout, presumptively, when it truly isn't gout.

And then, we are going to be telling these primary care docs that they will

be measured on quality for putting people on a med. And I just worry about people being put on urate-lowering therapy that don't actually have gout.

Now, hopefully, the fact that it does say that they have got to have hyperuricemia will help guard against that, but I still worry about that a little bit.

MEMBER ANNASWAMY: My comment is more of a clarification. The measure information form, like Helen was saying, clearly asks whether there has been evidence of unintended consequences.

So, the ACR is saying that at this point it doesn't apply. So, they have provided no information. But we are talking about hypotheticals. So, I am not even sure if we are voting on the information provided.

DR. PACE: So, the unintended consequences mostly comes into play with measures that have been tested and mostly implemented, because that is when you really

have the opportunity to see unintended consequences.

I think these are certainly things that are worth discussing and noting to the developer, that if they can look at any of this, because we do say, was there any evidence of it during testing. So, I think it is worth noting. And I think you should vote the way you think.

Yes?

DR. FITZGERALD: Yes, just to address the concerns. We do agree about the lifestyle recommendations. In our education, in our guidelines, all patients with gout, whether it is minor or severe or tophaceous, are advised and recommended that their physicians guide them and educate them on therapy.

For those patients who are not severe, lifestyle alone may be appropriate. For those patients who are severe with frequent attacks, lifestyle alone is likely

not to be effective and urate therapy will be needed for that.

And so, by looking at just the more severe patients, those patients who are with or without lifestyle likely are going to need urate-lowering therapy. And in those patients, you know, one of the main side effects is, as has been brought up here and discussed, is a gout attack. These are patients who are already defined as having frequent gout attacks.

In the short-term, increased risk of gout attacks for long-term reduction is a tradeoff. And again, hopefully with education, I mean, that can be moderated. These are patients who are having gout attacks. So, hopefully, they have therapies for their gout attacks, which would minimize potential harm.

The other harm is not treating and letting them go on to continue to have attacks and further progressive damage.

CO-CHAIR CHOU: If there are no other comments, I guess we are voting. Again, I think we are voting on whether we think it is usable for testing, but not for endorsement yet, but just whether we would think that they should go ahead and test this measure.

So, why don't we go ahead and do the vote?

MS. PHILLIPS: Now you are going to vote on usability and use. Your options are 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

You may begin voting now.

(Vote.)

DR. PACE: Okay, we have 21 responses, and 1 for high; 12 for moderate; 4 for low, and 4 for insufficient.

CO-CHAIR CHOU: So, I think that barely makes our 60-percent threshold with the combination of high and moderate. So, I think that means that they can move forward with testing in terms of the usability and use

issue.

Now do we do an overall vote now on the whole measure? I think now that what we do is get an overall vote, with all the considerations, the evidence, the priority, the gaps, the measure specification stuff, and the feasibility.

And again, I think we are not voting to endorse at this point. We are voting to move forward with testing.

Any final comments before we do this vote?

MEMBER VISCO: If this comes out as a "no," can it still be tested?

DR. PACE: So, if this comes out a "no," first of all, it goes out for public comment to see what the public comments say about it and, also, if there are issues that the developer wants to bring back to you.

But by no means, it wouldn't carry the NQF approval, and we should tell them what needs to be fixed to get NQF approval as a

trial measure.

So, you know, it really is a question whether any of the issues that were brought up are kind of fatal flaws for moving forward with testing.

CO-CHAIR CHOU: And can I just ask --

DR. PACE: Yes.

CO-CHAIR CHOU: So, really, the testing, it is just a resource thing. It is not going to have a clinical impact at this point. It is whether we think that it is worth the resources of the NQF and the partners to go forward with testing. Is that correct or are there other --

DR. PACE: Right. It really is whether this is a measure that is going to be useful and it bears further testing. The developer already thinks that. They have already invested a fair amount into specifying it to get to this point and submitting it.

And so, I think, you know, it

really is whether there is any fatal flaws that you think, you know, as you have said, it is not worth the resources to proceed with further testing.

So, we will correct this next slide for the next trial measure we have. So, disregard that it is -- well, let's just go to the last slide that is the yes or no.

Yes, yes. So, we will have a specific slide about whether you want to approve it as a trial measure. So, this is just a generic yes/no slide, and we will fix that on the break.

So, it is, basically, you will be voting on whether this measure should be approved as a trial measure, meaning you think it should proceed with testing.

Again, the testing would have to come back before this measure would ever be considered endorsed. And this, as well as any of the recommendations you make later, are things that will go out for public comment.

So, this is the first phase of the process.

So, why don't we go ahead? A yes would mean approval as a trial measure, and no would be not approved.

MS. PHILLIPS: Okay. We are voting on Measure 2550 for approval for trial use.

You may begin voting now.

(Vote.)

MS. PHILLIPS: We have 20 responses.

Okay, 14 say, yes, go ahead with the trial measure, and 7 say no.

CO-CHAIR CHOU: All right. So, I guess that passes and it meets our 60-percent threshold again.

And I think it is time for a break.

(Whereupon, the foregoing matter went off the record at 11:39 a.m. and went back on the record at 11:56 a.m.)

CHAIR CHOU: Just wanted to let

everyone know, we're going to talk about 2521 first before 2549. I think that it makes a little more sense clinically to talk about wether to measure before we talk about targets. I believe these are both trial measures. Is that correct?

MS. FRANKLIN: Yes. These are measures that will also be considered for cross measure approval. So they are not measures to be approved for ambility purposes.

CHAIR CHOU: And I'll turn it over to the developer, John, to give us a little, a brief overview.

DR. FITZGERALD: And thank you. So 2521 is titled, Gout: Serum Urate Monitoring. And the numerator statement is that patients measured with serum urate within 6 months after the date of a new or a change in dose onto urate lowering therapy prescription.

The denominator would be adults over 18 with established Gout initiating or

changing a dose of urate lowering therapy. The rationale for this is we want to look at patients who are given either a new prescription for urate lowering therapy or a change in dose. And we want a follow up urate measure.

The evidence shows that the majority of patients are given a prescription of uric acid and then, depending on the studies usually it's only about 20 percent of patients get a follow up urate level checked.

And so without a urate level being checked, you don't know if the patient's adherent, if the drug is effective and so you really need to connect the outcome of the urate lowering drug, which is the urate measurement. And if we have questions later about the data I can answer those.

CHAIR CHOU: Great. I think Steve was going to be the initial lead discussant and just provide an introduction.

DR. BROTMAN: Right. So, again,

this is item 2521 Gout: Serum Urate Monitoring. steward is the American College of Rheumatology. And a brief description of the measure once again is percentage of Gout patients who have started on or changed urate lowering therapy who's serum urate is measured within 6 months after dose change.

The numerator statement is that the patients with serum urate measured within 6 months after date of new or changed urate lowering therapy prescription. And the denominator is adults more than 18 years of age with established Gout initiating a changing dose of urate lowering therapy.

There are no exclusions. It's a process measure. It is another one of the electronic type of trial measures where the reliability and validity testing will be done at a later time.

And we'll start with a discussion of the evidence. The workgroup had a robust discussion of the evidence and I'm going to be

very brief because I want to get to a full discussion here.

But basically the discussion entailed that there's no direct evidence to support the proposed measure. Namely, there's no cited trials of uric acid monitoring versus no monitoring or treating to uric acid level targets or other strategies.

And the evidence is based largely on the association between uric acid levels and recurrent Gout. And given a lack of some of that evidence, it was hard for a lot of us to determine how high a priority this should be.

Some other workgroup comments. Some of us noted that while clinical guidelines were presented to support the measure, very little of the evidence directly addressed the impact of and the linkage between monitoring serum urate levels and improved outcomes.

Again, these measures are based on

the American College of Rheumatology guidelines in which some of their recommendations are level A but the majority, I believe are grade C recommendations. And that was noted in our final workgroup as well. And that other evidence would support that the measure focus is not available.

Other comments that there are not randomized controlled trials of uric acid monitoring versus no monitoring was mentioned before or treatment for uric acid therapies or other strategies. Although the developer noted that there are observational studies that patients with more monitoring have better uric acid outcomes.

And I believe they supplied some additional evidence yesterday that was forwarded to the committee, if I'm correct. And also the workgroup members were concerned that Gout is not always treated with uric acid lowering therapy.

And some of the members noted that

other treatments and patient education about Gout may also have a significant impact on outcomes. So I wanted to just go back to JD, if you had any comments before we discuss a little bit more of the evidence.

DR. DANIELS: Thank you, Steve. And I'll be brief here. I'll try to be like Johnny, may as well just quit scrambling and throw the ball. So what they did was they gave us some more information. We went through that last night.

And they kind of, I think, at least helped me try to identify the inclusion criteria a little bit. And so they wanted someone who was 18 or older, had an active diagnosis of Gout that does not end during the measurement period, one health encounter with a health care provider for Gout and a medication order for uric acid lowering therapy.

As far as validity, they used the ICD-9 codes and they quoted a Herald, which

was a big paper. And they basically thought they had a positive predictive value of 61 percent and that they thought that they would have a good way of kind of capturing this.

They also quoted a guy, this Jackson fellow, who basically the potential of unmet need for Gout diagnosis and treatment capture, recapture analysis of national administrative debt. I read that three times, had a headache. I'm unclear on what that adds to it and maybe it will edify us.

Other additional evidence they gave was that serum urate lowering is endorsed by a number of guidelines including UR, the Brit and the American guidelines. And the FDA use serum lowering data as the primary outcome for drug approval when the new drug came out and that they thought that there was improved patient outcomes that correlated with these lower levels.

And then it really wasn't referenced but it was basically sort of like

a comment about during the discussion at the workgroup level there was talk about how this has had some same of the issues as measuring hemoglobin A1c cholesterol and not really linking to outcome.

And basically their response was that there's no literature in the rheumatological arena to support this argument and that there are much more robust resources for the American heart than they do for Gout.

They don't think that they're going to be able to approve that. And they felt that it was unfair to compare the two. And that as basically what I got out of the last one.

CHAIR CHOU: Jonathan, did you want to respond?

DR. FITZGERALD: Sure. I can go through those points. So, again, there was more in common about this case of the patients. If patients are not on urate lowering therapy then they're not going to be

in this measure. So if someone's put on diet or lifestyle, there's not a requirement to check the urate.

What we're asking is if physicians or providers start a urate lowering therapy or change a dose then there's a measure. And the rationale for that is to look for the changes. It is true there's no monitoring trials but we are going off the data showing that urate levels are tightly correlated with outcomes.

And so that is important to follow urate levels. When we had our telemeeting, the teleconference, there was a lively discussion about serum urates and this gets to the next level also. But the value of serum urates and whether it was postulated that patients could just be put on urate therapy and not monitored and not have the serum urate monitored.

And the argument was in the atherosclerosis and with the lipid literature getting away from lipids, there was also

concerns about in diabetes where hemoglobin A1cs were presented as the target that then there was, you know, adverse events from people titrating down too low.

And a good bit of the conference had focused on that. And so we wanted to respond to that. You know, we think that these situations are quite different between, well, we're certainly trying to make an analogy that this is an important intermediary outcome and so there's some similarities.

There are important differences. Serum urate is tied with outcomes and these drugs, the urate lowering drugs work primarily through lowering the serum urate. So that's the rationale for monitoring.

Gout is not atherosclerosis. And there aren't the resources that are going to be put into it. And there have not been any trials. And it's unlikely that there would be trials randomizing groups of patients to whether they're going to be monitored with

their serum urate or not monitored.

We do have observational data looking at international studies. British and German large population samples showing that patients who aren't getting monitored tend to have more gouty attacks. Also have, when they do get monitored, they have higher urates.

And in contrast to the hemoglobin A1c, there's not a known risk for driving down urate below six, five or four. In fact, when the Pegloticase studies, they're really driving the urates down to undetectable levels. And we don't have adverse outcomes. We have the treatment goal.

So I think there are differences between the other intermediary outcomes. The other point that was being discussed was on validity of the Gout diagnosis.

So the primary study, looking at that, is the Arroll study looked in a VA database looking for the Gout diagnosis alone. All of our measures are Gout plus urate

lowering therapy. So it's not a direct comparison.

The positive predicted value was 61 percent, which was similar to other rheumatic conditions. It's not great. The concern with these studies and doing eMeasures is not so much the sensitivity but the specificity. You don't want to incorrectly put people into your measure. So the specificity is more important.

If the sensitivity is off, that's okay as long as you're looking at big samples. If you're missing people, as long as it's fairly random, that shouldn't be as harmful.

Because our measures are using the ICD-9 codes for Gout plus a prescription for a urate drug, it will increase the specificity. Now, admittedly the Jackson article is close to torture. It's very difficult to get through.

The point of that study is that they had done the same thing. They had used

a urate diagnosis to increase the specificity. Now they didn't do formal testing on that. But they said it would have been near 100 percent, which is sort of, I think, the author's opinion.

But there have been studies that, again, have looked at the ICD-9 codes. There have been other studies that have used a prescription for a "Gout-specific" medication. And those would be urate lowering therapies.

Other authors have used colchicine, which I would argue is not Gout-specific because there's others. We're looking at the urate lowering agents. The concern with, particularly allopurinol is that it can be used with patients with leukemia or lymphoma to prevent.

And so authors have excluded to leukemia or lymphoma and it has been suggested to us that we consider excluding those. We haven't specified that but we can certainly test that when we go into the testing phase.

The rationale for not excluding is because we're not using allopurinol as a urate lowering therapy or a Gout-specific medicine as a criteria for the diagnosis. We're using it joined with the Gout diagnosis.

So it would really be a small number of patients who were Gout and lymphoma or leukemia and on urate lowering drugs. So we hadn't specified it that way but we can certainly look into that when we're doing the testing. So I think those are the various questions that were raised.

CHAIR CHOU: Questions for either John or lead discussants or comments? Yes.

DR. MATUSZAK: So, just hypothetically speaking, if this is your quality measure and you've got a patient that you started onto urate lowering therapy because they've had multiple gouty attacks and they started off with the serum uric acid level of eight and they come back to you to get their test and it's four or five months

later and it still shows eight but they haven't had any gouty attacks, are you treating the number or are you treating the gouty attacks? And, you know, is actually getting that number improving the outcome for the patient?

DR. FITZGERALD: So rather than go with a specific patient, I'd go with what the data shows for groups of patients. And if you looked at groups of patients, patients whose uric acid remained at eight would have a higher proportion and the data would say it's about a 50 percent higher risk of having a Gout attack versus those patients who it was reduced down to six.

And if you go up higher, it can go up for patients who their uric acid that was in the range of ten or higher, the odds very sure would be 2.5, based on some other studies. So there is evidence to argue that you should be treating the number in that specific sort of case instance. I think it's

best to think about what the data is.

CHAIR CHOU: Yes?

DR. GHOGAWALA: This is a feasibility question. But it relates to something that we experience in our health center. Many of times we have patients who are seen by a rheumatologist or an endocrinologist or a specialist started on medical treatment and then referred back to their primary care physician.

And the question that I have here is, and we're under enormous pressure to do that, as I think a lot of us are. How would we be able to, from a feasibility standpoint, measure whether the physician was doing a follow up study or not.

Because at least in our case, many of these doctors are not using the same electronic health records as we are. They are primary care doctors.

DR. FITZGERALD: The sites we're looking at have unified health records, you

know, for the eMeasures you're really reliant on testing and applying those in a situation where you could capture data.

The patient's going to outside labs or paper charts, that will create problems. But that's true of any eMeasure. And from a practical point of view, whoever, you know, as long as the uric acid was drawn, it doesn't matter who drew it if the lab measurement was there after the prescription or dose change.

DR. GHOGAWALA: Understood. Thank you. From a quality perspective, if the measure of quality here is the initial prescriber and if that initial prescriber has no control over beyond a recommendation, I think this may be more common if you look at this from a natural scale than you realize.

And in fact our scenario here is not that people that have paper charts. Most people have electronic charts but it's just different and not compatible with our central

health care system in all cases. And we are under pressure for, you know, independent medical groups to send those patients back to us.

And in fact I would think that Gout would be an example where you wouldn't necessarily keep it in a tertiary health center. You would send it back.

DR. FITZGERALD: So that concern would be, for example, to the rheumatologist who might start a urate lowering therapy. Personally I would have a higher expectation about that prescribing physician following up on lab measures for a new drug they're starting before signing back. So I think the responsibility would still be there.

DR. ANNASWAMY: So from the comments and the evidence provided, it seems like there was no direct evidence to suggest that monitoring a change in uric level leads to better outcomes. So there is indirect evidence. And also the six month project,

also there is no direct evidence.

To the question of clarification here about whether it's the highest it can, what's the highest ceiling that we can potentially raise this evidence because there is no direct evidence to this recommendation.

DR. PACE: Right. And maybe we can go to that slide and let's look at our algorithm here. So the focus of this measure is follow up serum testing after prescribing the drug or changing the drug.

And the evidence that was provided is about, I was just looking at this, it's not directly focused on that topic. So basically you're saying there's insufficient evidence for what they're measuring in the topic.

DR. ANNASWAMY: There's indirect evidence.

DR. PACE: Indirect evidence.

DR. ANNASWAMY: So there's indirect evidence going in a particular academy?

DR. PACE: Right. Good question. Go ahead. Do you want to?

CHAIR CHOU: Yes. I mean, I was hearing people saying, you know, indirectness is something we deal with. And grade deals with that as well and so does the task force and others. And there's many different types of indirectness, of course.

So what we're talking about here is that we basically have correlational studies but studies of the actual intervention, which in this case is uric acid monitoring.

And I think that there's, you know, if you use grade you basically, you know, look at the overall body of evidence and ding it for indirectness and you can ding it lower or less severely depending on how indirect we think it is.

So it requires some, you know, subjective, you know, judgement there. You know, the degree of indirectness, essentially

here. But I think it would be tough to get above a moderate. I mean, to me at least it would be tough to get above moderate without a study, you know, showing that monitoring actually improves outcomes.

DR. PACE: Right. And that's consistent with what our guidance is and what the evidence task force had identified. So the evidence is not directly about what's being measured and, you know, the highest it would be is moderate. But even, I don't know if they gave a grade in this. But someone was talking about C level, which is probably expert consensus, so.

CHAIR CHOU: Yes. I was going to say that John mentioned some of these cohort studies or whatever that looked at monitoring versus no monitoring. That's direct evidence.

I mean, they're observational but it still, you know, would more directly address this so it would be nice to see that because I too am troubled with the lack of

direct evidence at least in what's been provided so far.

I also have a question about what about people who start, you know? So two issues, one is that people with acute Gout often have, you know, normal uric acid levels.

And what's the rationale for, I don't know, I mean, or they have, or even with, you know chronic Gout they have uric acid levels maybe of less than six. And what's the utility of monitoring in those situations.

Also just to note that again in the febuxostat trial, they didn't monitor the -- I mean, monitoring may have been part of the protocol but the way they treated was just, you just treat. If somebody meets these criteria you treat. And they were able to show that they had a lot less Gout attacks on either dose of the febuxostat.

And so at least if you're going to go by the trial data, it seems to me that, you

know, whether treating is what we have the clearest evidence that if you start somebody on one of these uric acid lowering therapies that you do impact Gout outcomes.

DR. FITZGERALD: Is used as a monitoring versus nonmonitoring. It was really designed to show the efficacy of their drug.

So, some of the other points I wanted to get back. So, the serum uric acid, Lesson 6, I think we'll defer some of the level discussions to the next point, where we do specify where we think a target should be.

This really is, if the provider has started or changed the drug, it means that they are unhappy with where their current gout management is and they're trying to move to a new state. And uric acid levels change within 14 days and what we're asking is that there be a follow-up, so there's information about that new state.

So, regardless of what the level

was, the provider thought that the urate needed to be managed or changed. And, so, that's the rationale for that.

And, then, the final point about the direct and indirect, some of this measure was designed for phase validity. Construct is the provider is writing a prescription for a change in urate therapy and the natural follow-up would be, then, to measure that urate level to see if the change in the therapy was effective.

And, then, it had been previously mentioned about supporting measures that had precedent and, in preparing for this, I went back and was reviewing the data. And there was a group that proposed gout quality care indicators back in 2004. It was Ted Miklus, Ken Saag and Kathy Mclain group.

And they came up with a measure, really matched a lot of our measures. But theirs was a gout patient is given a prescription and, at the time, it was just the

xanthine oxidase inhibitor was just the primary indicator.

Then a serum uric level should be checked at least once during the first six months of continued use, because periodic serum uric measurements are required for appropriate dosage estimates of those xanthine oxidase inhibitor for escalations or reductions. So, there's been some precedent for this type of measure.

DR. PACE: So, I want to just go back to the question that was posed to me, in terms of where this falls in the algorithm. What's been provided is not necessarily even graded. So, it's either going to be moderate or insufficient. And, then, the question would be whether it's insufficient with, you know, justification for an exception.

But even with the moderate, there should be a grade of the evidence and it should be a fairly high grade or strong recommendation. So, from what I'm hearing you

talking about, it doesn't even seem like it fits in that box. But that's certainly your judgment.

CHAIR TEMPLETON: Yes. Go ahead, Steve?

DR BROTMAN: Yes. Can I just ask for a clarification? There's two contrasts. One is going on clinical practice guidelines, which does look at some evidence and, depending upon what that is, it's graded.

But we still have a hybrid here, if I'm correct, where they provided some other evidence and, at that point, is it our job to look at the quality, consistency and quality of that evidence and grade it accordingly?

DR. PACE: Well, we basically don't think that's a Committee job. The Committees aren't constructed to do systematic reviews of the evidence. And, typically, it's not a measure developer job and, so, we ask the developers to use graded evidence.

However, you know, when you're in

this area where maybe there isn't a systematic review or they've submitted kind of individual studies, it is going to fall to your judgment, you know. The question that always comes up, when we have individual studies submitted is, is that really representative of the body of evidence or is it selected study?

We just don't know, because it wasn't a systematic review and approach to gathering all the evidence. But I think Dr. Chou can guide us more on that.

DR. CHOU: Yes. I mean I was going to say that. But it seems to me that we're both in Box 6 and Box 7, that there is a guideline and the evidence isn't quite what we're looking for. And, so, we're also in Box 7, which is not really captured in this algorithm, being in both places at the same time.

You know, Box 7, 8 and 9 do allow you consider other stuff. So, I mean I think it's possible. I think we need to be cautious

about that because, you know, without a systematic review, you don't know what's missing or whether it's selective, you know, how they pick the studies and all that kind of stuff.

I guess, Steve, you're going to follow up?

DR. VENTURA: It seems to me that several of the other measures would face the same thing, that there's no direct evidence that the process measure itself improves the outcome. There is a lot of indirect evidence about the steps in the process.

DR. CHOU: Right.

DR. VENTURA: So, we could still use Box 7, couldn't we --

DR. CHOU: I think so.

DR. VENTURA: -- if it's still happening?

DR. CHOU: Yes. Yes.

DR BROTMAN: So, then, does the question become is there a sufficient amount

of evidence or insufficient amount of evidence? It almost seems that way sometimes.

DR. CHOU: Yes and I think that's what the discussion right now is. If we don't think the evidence is there, we don't move forward. I mean we kind of stop there. And, so, I think we want to have a little bit more discussion here, before we go on.

I just wanted to just comment on the positive predictive values and stuff. I mean because this is relevant to the first measure we looked at, actually. That's pretty terrible. That means that one out of every three people who has an ICB9 Code of gout doesn't actually have gout. I mean that's pretty bad.

And it may be the same for RA and OA, but it just means it's bad for all of these conditions. So, that's concerning to me. I mean I hope that that's something that's considered in the testing also, to really look at whether we're really actually

getting people with gout. I mean everyone's brought this issue up previously.

DR. FITZGERALD: Yes. Again, that would be sort of the floor, because we're looking at all their measures are gout plus a urate lowering. There could still be errors there.

But the Jackson study that was provided, again, hadn't done a formal testing. But they had argued that the specificity on it to be very high. They claimed a higher specificity.

DR. CHOU: Yes. Yes. I tried to look at that study, too. It's hard to really determine. And it's not clear to me how, just because somebody's being prescribed it, how suddenly the diagnosis becomes.

It just means somebody's prescribing the drug because, if they already gave an incorrect diagnosis and are prescribing the drug, it doesn't make the diagnosis any more correct. So, that's why I

have some questions about that study.

But, again, I think this is important from a reliability kind of issue. That's problematic. So, other comments or concerns about the evidence?

DR. DANIELS: Just one quick clarification.

DR. CHOU: Yes.

DR. DANIELS: I might have gotten this wrong. Basically, what I'm getting is that Jackson reports that the addition of allopurinol and colchicine improves the sensitivity to 84 percent?

DR. CHOU: The sensitivity is not the part it predicted. The part of the predictive value is how many people you say have the condition that actually have it. The sensitivity is how many people with the condition you can identify.

And, so, the part of the predictive value is really what we're looking at, in terms of the accuracy of our ability to

identify the denominator.

DR. DANIELS: I'm sorry. I was talking about the inclination to move closer to 100 percent S values.

DR. FITZGERALD: I think I also pasted a comment. So, they reported the sensitivity and they fully tested for the sensitivity and that had gone up to 84 percent with their methodology. They had this separate comment and I pasted it, just so I was using their words and not an interpretation and they described that the specificity was increased.

DR. DANIELS: I see what you're saying. So, as far as the reimbursement is dependent on submitting returns, 100 accurate isn't likely. They didn't actually get it, then.

DR. FITZGERALD: It's a nonspecific statement.

DR. DANIELS: Yes.

DR. FITZGERALD: But they were

arguing for a higher specificity.

DR. JARRETT: Getting to the issue of, you know, this is really bordering on whether or not the population that we're dealing with is not a homogeneous population. So, there will be groups of patients where, perhaps, you know, lowering them from 7 to 6.5 may have no clinical significance.

However, as a process measure, if we don't kind of have broad stroke with this, we're going to miss that large group where really truly the number and the measuring of the number really has impact.

So, I think we have to look at the broader population, realizing that we're not going to be able to slice and dice it the way we ideally would like to and maybe ten years from now we can. But we'll have to take a broader view and say, yes, it should be measured. There is enough evidence.

Of course, there is a large segment that that measurement and making sure

the number then eventually does come down to the appropriate number really does make a difference.

CHAIR TEMPLETON: Yes?

DR. DODGE: This strikes me as similar to the examples you gave earlier about insufficient evidence but with a exception, because you've already set the bar as the criteria for inclusion are that you have initiated therapy or changed therapy.

Then the urate is in the discussion as important. And it would seem completely obvious that you would want to see what that effect of the intervention has been. And I don't know. It would be so limited to try to find a study that proves that just that measure is enough.

But it is a prerequisite I think to seeing if the other things like a serum uric target are actually effective. So, this is one of those ones where I don't know as the burden of evidence for me had to be very, very

high because it just seems like a very straightforward accountability to the decision that is an inclusion criteria.

DR. CHOU: I just wanted to respond to that a little bit. I was one of the people that brought up the issues with lipid treatment and diabetes treatments, where there's a clear correlation with A1c levels and bad outcomes. There's a clear correlation with lipid levels and cardiovascular outcomes.

But you can take two drugs that have the same lipid lowering affects and statins reduce heart attacks and other drugs don't. And same thing with diabetes that, if you take Metformin, you can reduce, you know, events. If you take Rosiglitazone or whatever, it doesn't. They have the same exact effect on the A1c level.

So, there's more than one type of indirectness here. So, now, not only are we looking at indirect evidence in the sense that we're just looking at correlational

observations, we're also looking at intermediate outcome.

And this is what I was talking about is that there is degrees of indirectness. And, again, I just think that this is something we have to be careful about, because many groups have gone down that path before.

I'm certainly not arguing that their path of physiology is the same. I don't know that, you know, maybe allopurinol and probenecid and all these other drugs have exactly the same mechanism and it's all uric acid. The evidence I don't believe is very, you know, sound on that. So, I just want to make that point. Yes?

DR. MATUSZAK: And I think if I understand, our function here is not to necessarily say that, I mean something might be very, very good to do in clinical practice, but if the evidence doesn't support us measuring it as a quality outcome, if the

evidence doesn't support that, then I don't know that we necessarily have to say.

I mean it might still be a great thing to do clinically but I don't know that we need to endorse or support measures that don't have the evidence behind the process.

DR. FITZGERALD: If I may comment, as you pointed out, it seems incredibly obvious to check a urate after you've started a urate drug. And that doesn't happen. In only 20 percent of patients who were given a prescription get uric acid checked in the next year.

It would be akin to finding someone who has hypertension starting a anti-hypertensive and, then, not checking a blood pressure measurement. You're not going to do a trial on that and I think we're likely not going to see a trial on whether you should check a uric acid after starting or the benefits of or the timeframe or the window.

So, I think we're going to be

challenged with that. So, I think a lot of this is dependent on phase validity and documentation that there are gaps in what would seem to be obvious care.

There are fewer treatment alternatives for lowering urate therapy. Uric is used only about ten percent of patients. So, it's really primarily allopurinol and febuxostat has been growing. So, there are only a couple treatment options.

The data suggests, through a randomized trial, when they did the allopurinol versus febuxostat, had similar outcomes as far as gout. And that's not saying that that explains everything. But the mechanism, when you look at the data, is patients who respond to the drug with urate lowering are less likely to have attacks.

So, it does seem to be really an intermediary and the way these drugs work is by lowering the levels down below solubility. And we've seen that with ultrasound studies,

when patients are put on the medications, that there is resolution of the findings of the crystals in the joints, again, indirect.

MR. SCHUNA: I'm not sure it's akin to blood pressure lowering and hypertensive treatment in that you can start uric acid lowering therapy and the patient no longer has gout attacks or has less frequent gout attacks. Maybe that's the goal for the patient in many cases.

DR. FITZGERALD: We did not tie indications into this measure. Simplified, again, conceptually, we're thinking these are patients who are having frequent attacks. The problems that are reported with adherence and with patients who respond to the drugs have dose titration, Perez Ruiz presented the article that most patients are on 300, but only a proportion of those meet what would be considered the target for having levels less than six.

And it's because monitoring is not

being done or is not being done in a timely manner or there's not responses to monitoring. So, there's several problems in the process of care between getting the test, responding to the test, titrating the drug.

And what we're trying to do here is get people on that first step as far as monitoring the change that was initiated.

DR. YAZDANY: Can I just make one more point?

CHAIR TEMPLETON: Yes, go ahead.

DR. YAZDANY: So, I think with the pathophysiology of gout, although it wasn't discussed in our submission materials, is very relevant here. The question is how good of a surrogate intermediate outcome is uric acid?

And I think, unlike many of the other parallels that people are drawing, uric acid is actually found in the tophus. It's actually found in the joint that's having a gout attack with incredibly high sensitivity.

We have sort of a human model of

gout where, if somebody has cancer and we give them chemotherapy, cells release uric acid, so you get a huge load of uric acid and you get an acute gout flare in the setting of chemotherapy and we have prophylaxis.

So, gout is not only at the crime scene. It's found, you know, consistently with high sensitivity in the joint. So, unlike things like say cholesterol or even with hemoglobin A1c, where the distance between an outcome, say a myocardial infarction and intermediate outcome are far, I would argue that, in gout, the intermediate outcome and the final outcome are actually very closely linked.

And that's why there won't be a trial of monitoring uric acid or not monitoring uric acid because it probably would be challenging for an IOB to approve that.

DR. DANIELS: I'm going to say something to that and I'm going to use that Perez Ruiz article that they had. And one

thing that I found was, you know, only three percent of the patients with gout received doses over 300. However, and this is where it's kind of saying all the way to below six, which some people would say a low level, that they only needed 370 milligrams.

So, you know, there's obviously some other things happening here besides just the uric acid level. You know, there are some things that are left out that we don't really understand because that tells you there. It's almost to what Art just said that, you know, there's going to be a certain number of patients that you're going to put on medicine and it lowers and they kind of do okay.

So, you know, it looks like we're kind of under-treating but it's still, at least according to this study, not that bad.

DR. FITZGERALD: That was the average dose requirement and I have to go back and look at the proportion of patients who had met target, even if we just assume an average

dose, it would imply somewhere around half weren't reaching the target. I'd have to go back though.

DR. CHOU: Karen?

DR. PACE: Yes. I just wanted to emphasize what Dr. Chou said, in terms of kind of the whole causal pathway, one of the things that we do ask the developers to do that I see wasn't in these forms is to lay that out, the pathway between what they're suggesting be measured and the ultimate health outcome.

So, the way I understand this is that you're talking about measuring uric acid levels that should result in looking at the uric acid level and what's the connection of that uric acid level to decreased gouty attacks and symptoms and, you know, if you lay out that causal pathway, where do you actually have the evidence?

So, I think we've been talking about the evidence of giving these meds and it lowering the uric acid level. But the

question that seems to be coming up, then, what's the relationship between that level and the attacks.

DR. CHOU: Yes. I was going to say I mean there's a lot of assumptions that physicians will act on the level and that patients will actually take the medicines and all these other things. I mean there's a whole slew of other things that would have to happen, which is why it would be nice for someone to actually study it to see what the impacts are. But go ahead.

DR. FITZGERALD: I think this measure is as subjectives are modest, because we're trying to get physicians to do the first step, as far as checking the level. And, as far as laying out the causal pathway, I'll just highlight the studies that were submitted.

The Soji article is the one and that's in here showing the uric acid levels and proportion of patients who have attacks.

There's a Perez Ruiz article who describes the rate of tophi reduction and the velocity of tophi reduction is highly correlated with uric acid levels.

As mentioned, Tio and colleagues have purported resolution of crystals with improvement in the management of hyperuricemia. So, those are scattered through here.

DR. CHOU: And I'm still not entirely clear about the rationale for monitoring somebody who has a uric acid of six, say, to start with. What would be the rationale for that?

DR. FITZGERALD: So, that might be better addressed in a different section. In the guidelines, if someone has a level of six, we would be happy with where that is and there would not be a dose change and they would fall into this. I mean, if their symptoms are doing well --

DR. CHOU: What I'm saying is

someone has six when they have their first gout attack or their second gout attack. You start them on uric acid lowering therapy. Why would you need to check their uric acid level again?

DR. FITZGERALD: So, again, we're coming up with patients who are really at the margins of what we're describing in these measures. There are several issues that I can address with them.

DR. CHOU: About a third of patients with gout have uric acid levels that are around six. I mean they're not very high. It is my understanding that there's quite a few patients, actually, that don't have highly elevated uric acid levels.

DR. FITZGERALD: So, again, if we're talking about patients with frequent attacks, the majority of patients will have levels of eight or ten, so, significant hyperuricemia. It's certainly possible to have gout patients with uric acids of six or

even lower. Those are really the exception.

There is also the known phenomenon that, during the attack, the uric acid level can be dropped and artificially low. So, there are several issues in there. I mean correct me if I'm misunderstanding the question, but I'm not sure that they're on top of the measure.

DR. ANNASWAMY: I think I can add to what Roger is saying. The denominator is essentially patients who had a new drug initiated or change in drug. So, among those, perhaps there is reason to exclude those that do not have high uric acid levels.

They may not need monitoring of their levels as much as others may. And the evidence does not support such a robust need for those patients and they have not been excluded in this denominator.

DR. FITZGERALD: I think that's a very good suggestion. We certainly weren't looking toward that in excluding patients who

had levels lower than someone would act on, because you're right. At those points, if someone has a level of five and they're still having attacks and it's determined that they need more urate therapy, it's not necessary to check that.

I do think that is a small minority of patients. Those are more likely ones that are getting to the rheumatologist. So, you know, 20 percent. And then, for the rheumatologist, that's an unusual patient still. That would be two percent of that. So, I think it's a small number but I think it's a very reasonable proposition.

DR. CHOU: Other comments? I think it's time for us to take a vote on the evidence. So, again, this is a trial measure. So, we'd be kind of voting whether there's enough evidence to move forward with testing. And let's go ahead.

MS. PHILLIPS: We've got one for high, two for moderate, three for low, four

for insufficient, four for insufficient with exception and five for insufficient evidence. We're voting on Measure 2521, points to measure and report. Starting now.

We have 21 responses: five for moderate; five for low; six for insufficient evidence with exception; and five for insufficient evidence.

DR. CHOU: So, do the two and four go together, is that how it works?

CHAIR TEMPLETON: Yes.

DR. CHOU: So, it still doesn't quite meet the 60 percent threshold. So, it's 40 to 60 percent. Can you remind me what happens here? Does this mean we wait for a public comment to come back and then reconsider?

MS. FRANKLIN: Yes. I think at this point it means that it would continue on and, of course, we've taken public comments on the measure and the Committee reconsiders the measure after the public comment period.

DR. CHOU: All right. So, the next area is opportunity for improvement. I think this is related to a lot of the stuff we've already talked about. So --

MS. FRANKLIN: We'll have to pause it for a public comment.

DR. CHOU: Oh, okay. Sorry. Pause for a public comment.

MS. STREETER: I'm just going to try to keep us on a schedule here. We'll take a quick break for public comment and then we'll have our lunch break and continue on to a performance step after that. Operator, at this time, can you open the lines and see if we have any public comments?

OPERATOR: If you have a public comment, please press 4-1 on the phone keypad. And there are no comments at this time.

MS. STREETER: Okay.

DR. CHOU: I think we get to break for lunch. Where is lunch?

MS. STREETER: Lunch is

(Whereupon, the foregoing matter went off the record at 12:49 p.m. and went back on the record at 1:16 p.m.)

CHAIR CHOU: It's 1:15, so we're going to reconvene here, try to get back on schedule.

So we're still discussing Measure 2521, serum urate monitoring. We have just finished the evidence discussion and we heard public comments, of which there were none. So I think we can move on. Do we need to -- okay.

So we're now at the kind of opportunity for improvement piece. Maybe, James or Steve, if there are additional comments. I think we've already all heard much of the evidence about potential performance gaps, but if there's anything that you wanted to add here.

DR. BROTMAN: I don't have anything. I think we've had the discussion mostly.

CHAIR CHOU: Okay. And, JD said the same.

So I think it's time for us to vote on whether we think there's opportunity for improvement here.

MS. PHILLIPS: Okay. We're voting on 1B, performance gap. Your options are one for high, 2 for moderate, three for low and four for insufficient. Voting begins now.

We have 20 responses. Has everyone voted who's going to vote?

Still at 20.

We may need to redo this. There we go. Twenty-one. Thank you.

Okay. The results are 2 for high, 11 for moderate, 2 for low and 6 for insufficient.

CHAIR CHOU: I think we're right at our cutoff with moderate and high, right at 60 percent. So I think we move on.

The next area is priority. Again, any additional comments from Steve or JD about

priority that we haven't discussed previously?

DR. BROTMAN: Priority is still similar in terms of the work group discussion that gout may be a high priority, but not sure if the lack of monitoring represents a high priority gap and a high priority.

CHAIR CHOU: Open it up for any other comments?

(No audible response.)

CHAIR CHOU: I think it's time for a vote here.

MS. PHILLIPS: Okay. Voting on priority. Your options are one for high, two for moderate, three for low and four for insufficient. The voting begins now.

We are at 19 responses, so can everybody try again?

There we go. Twenty-one. Thank you.

All right. Zero for high, eight for moderate, six for low and four for insufficient.

CHAIR CHOU: All right. So I think this is the first time we haven't met at least the -- is that -- that's below the 40 percent cutoff, right?

MS. PHILLIPS: It's below.

CHAIR CHOU: So what do we do here?

MS. FRANKLIN: At this time that would mean that this is a must pass criteria, so the measure would not continue forward for consideration. However, this will also go out for public comment and the Committee can react to comments received from the public after the period is closed.

CHAIR CHOU: Do we take public comments now or do we --

MS. FRANKLIN: No.

CHAIR CHOU: -- wait?

MS. FRANKLIN: No.

CHAIR CHOU: No? Okay. So I think we are -- we don't go any further, so maybe we move on to the next measure at this

point. So that's 2550 -- no, not 2550, 2549. This one is of course related to the prior measure. Instead of being about monitoring, it's about targets.

I guess I'm not quite sure how to proceed here. I mean, if we're not going to suggest monitoring, I'm not sure how much sense -- I mean, if we're not going to move forward with that one, do we still want to discuss this one?

MS. PHILLIPS: Yes, we still have to discuss this one --

CHAIR CHOU: Okay.

MS. PHILLIPS: -- as a stand-alone measure.

CHAIR CHOU: Okay. So, John?

DR. FITZGERALD: And I would argue that -- I mean, this is separate and we designed it separately, so this was not meant to be conditional on the prior one. And the rationale for this one is quite different than the last one also.

So this is our serum urate target. The brief description is the percentage of patients with a gout diagnosis who have been treated with a urate lowering therapy for at least 12 months, that the serum urate checked is at least once yearly and with the most recent result being less than 6.8 milligrams per deciliter.

And the numerator statement is that adults, patients 18 years or older, in whom a serum urate level has been checked at least once yearly with the most recent being less than 6.8 milligrams per deciliter. The denominator is the patients with a gout diagnosis who have been treated with a urate-lowering therapy for at least 12 months. And we had no exclusions.

CHAIR CHOU: Steve, do you want to give an overview?

DR. BROTMAN: Well, I think that was a good overview of the measure title for 2549 gout serum urate target as well as the

developer rationale and the numerator and denominator statement. This is a process measure, another EHR measure where this is on a trial basis where their reliability and validity testing is not done yet. And they state that they will be doing it later and submitting for full NQF endorsement at a later time.

Just to go onto the evidence section, similar types of discussion, but a little bit more related to the targets. Related to the evidence is indirect, but the association -- it's basically an association between uric acid levels and gout, but no studies cite compared effects of targeting of less than 6.8 versus other targets. And other comments about they would rank the evidence as low or possibly insufficient.

The work group noted that the clinical guidelines are presented to support the measure. It's indirect based on that association of the levels and the gout attacks

and rather than the impact and the linkage between the monitoring of the serum urate with a target of 6.8 or less -- or less than 6.8 versus other targets in improved incomes.

So work group questioned whether the required quality and consistency of any evidence that had been put forward other than the guidelines had been met and noted that the evidence presented again was level C grade.

JD, did you want to chime in with anything?

DR. DANIELS: Yes, stuff that they gave us on top of that was basically they kind of talked about the Halpern study. Again, it's one of these things that doesn't exactly kind of match what they're measuring. It's the non-compliance of urate lowering drug.

And then they -- I'm confused about this, because they're saying we're doing -- ICD-10 will be used to identify patients with gouty attacks. So I didn't know if that's something -- they're going to use this

to actually collect their data for them? Is that what they're -- it doesn't look like there's any new data that they're giving us, but I don't know if they're saying that they're going to use -- if we say it's okay, that they're going to use this process to collect their data. That I guess would be my --

DR. FITZGERALD: So, sorry, I think there's some confusion on that. This form was -- I structured this form based on the question feedback we had received from the work group.

DR. DANIELS: Okay.

DR. FITZGERALD: And this question fell under here about how gout attacks were going to be defined. I think that question actually relates to the urate lowering therapy indication. So this answer -- the question and answer should probably be in another section. So we could disregard that here. Gout attack is not part of this measure.

DR. DANIELS: Okay. I'm just --

DR. FITZGERALD: Yes, sorry for that. Yes, I understand the confusion and I apologize.

DR. DANIELS: Okay. That's all right. A lot of this overlaps. And really nothing else.

DR. BROTMAN: And I just wanted to add that there was some discussion in the work group that this is possibly just a check box measure and just capturing a snapshot in time that may not actually reflect the compliance of someone on ULT.

CHAIR CHOU: So, John, do you want to respond to some of their comments?

DR. FITZGERALD: Sure. So there's a lot in there to respond to. A lot of this is related to some of the discussions we've had before. There is not evidence with a, for example, randomized trial looking at 6.8 versus 7 versus 6, and we're not going to have that. The data to support this measure is

that serum urate levels that are brought down lead to improved outcomes. And there have been the febuxostat randomized control trials looking at that showing improvements in patient outcomes associated with improvements in the serum urates.

I think the best evidence, again indirectly, is the Soji article, And I put a picture of the figure in the responses that we provided is the relationship between serum uric acid and future gouty attacks. The authors themselves had specified that at levels above 6 milligrams per deciliter there is increased attacks. In the 4,000-patient British and 3,000-patient German studies higher urate levels again were associated, and that's where I gave you the numbers before. People in the serum urate 8 had a 1.5 odds ratio of having more frequent attacks. Serum urate levels of 10 had a 2.5.

So when you look at Soji, even though -- I mean, they did pick six, and other

people who have picked six milligrams per deciliter includes the ACR, the gout guidelines, which was based on RAND/UCLA methodology. The EULAR guidelines, the British Society of Rheumatologists, the Dutch General Practitioners have all chosen six as well, some of them even choosing five for certain indications.

When you look at the data, even the Soji, even though it's somewhat sigmoidal, it is fairly linear. The Perez Ruiz study, that's also looking at velocity of tophi reduction, which is tightly correlated with serum urate levels, is also fairly linear. So it's admittedly very hard to pick a level.

Now, we had considered the six milligrams per deciliter because there was so much consensus on that, however, that would then penalize anybody who was getting close or close enough and we thought, again because it's a quality measure, we ought to pick a higher level, to be a little bit more

permissive so as not to ding someone who gets a uric acid of 6.4 and is doing well.

So we picked a higher level. We didn't want to pick an arbitrary level. And so 6.8 milligrams per deciliter is the solubility level for serum urate. So that level has been picked. That level has been cited frequently, is sometimes used as the definition. Seven is also sometimes used as the definition for hyperuricemia. We do plan during the testing phase to look at different thresholds to see if that's going to make a big difference in our outcomes.

And I think those were most of the questions. Were there --

CHAIR CHOU: I'll open it up to the rest of the panel. Other questions or comments about the evidence here?

MS. DAVIS: This seems to me to be an intermediate outcome in a way. I mean, it's actually a level and not a process measure. Is that accurate?

And then just a comment. I'm wondering if you just collect the data on the actual level on every one and then do some comparative analysis on it rather than setting it at 6.8 with a yes or no. Would give you more information.

CHAIR CHOU: Go ahead, John.

DR. FITZGERALD: I mean, we'll certainly look at that. We weren't aware of to be determined proposal as far as level. So we felt we needed to pick a level and now we just wanted to give you the background on how the 6.8 was picked.

CHAIR CHOU: Can I ask? So one of my questions is what's the rationale that -- if you've checked say once and their uric acid is 6.5 and they don't have any more gout attacks, what's the rationale for requiring a yearly level?

DR. FITZGERALD: So this -- yes, this was conceptualized for getting people started, so for the incipient year. I guess

as stated, it's continuing years if patients became -- if they're doing well. Lost adherence, levels would go up and they'd be likely to have attacks again. So --

CHAIR CHOU: Yes, but I guess clinically if they weren't having attacks why would there be any reason to recheck again once you've gotten their level down?

DR. FITZGERALD: I can't -- the only data I can cite for that is there have been trials looking at stopping urate lowering therapy, and even patients who've been well controlled after many years will typically flare after stopping therapy. So I think those were some of the earlier allopurinol studies. So I think there is value in continuing to follow.

CHAIR CHOU: Yes?

DR. GHOGAWALA: This is going to be just ignorance on my part, but is there any need to monitor any blood work for somebody who's say on urate lowering therapy for say

five years and is stable? Is there any other reason to be measuring blood work on them for that purpose? I just don't know.

DR. FITZGERALD: Yes, there's some safety monitoring recommendations as far as LFTs, renal function and, to a lesser extent, CBC.

CHAIR CHOU: Other questions? Yes?

DR. ANNASWAMY: I have a question about 6.8 versus a range. If less than six is associated with good outcomes and greater than seven is associated with bad outcomes, and you're looking for a snapshot and there is variability in uric acid levels day-to-day, time-to-time, and you have a range that is acceptable, perhaps a range is what you should be looking at given those variabilities.

DR. FITZGERALD: Well, we've said less than 6.8, so our range is 0 to 6.8. If we said less than seven, then we'd be -- I mean, some of this we will look at with the

testing data to see if there is meaningful differences. But again, we're picking -- our objective here was to pick a threshold, if the guideline is to try and to do something else. But even the range really is a threshold still.

CHAIR CHOU: I have another question just about -- again, this is one where I think that people with tophaceous gout and with erosions -- I would view them a little bit differently where the target or achieving a target may be clinically much more important than somebody who, for example, has a couple of attacks that are easily treated and you have them on uric acid lowering therapy and they're not having any more attacks. Why do you need to monitor those patients with a yearly uric acid? And so I just wonder if there was any consideration to focusing on people with kind of more severe gout.

DR. FITZGERALD: So in developing

this we tried to stay away from the indications and leave those as implied is that the urate lowering therapy has been prescribed for this patient, and therefore there are underlying indications. We hadn't specified the indications like we did in the first measure. Sorry, the rest of your question?

CHAIR CHOU: I think it was just --

DR. FITZGERALD: Oh, yes, for the patient that --

CHAIR CHOU: Yes.

DR. FITZGERALD: Yes, so urate levels -- well, urate levels are associated with tophus reduction, so, yes, clearly in those patients. And if it's aggressive tophus you want even lower levels. Urate levels are associated with recurrence of attacks. Again, for the patient at the edge or the margin, I have less data to support arguing that. We could try and again look at that and define some tighter -- we could put the indications

back in like we did on the other measure.

CHAIR TEMPLETON: And this is Kim. If I could ask a question, please?

So, I understand that the uric acid is the easy thing to measure, the easy thing to find in electronic health records, but wouldn't a more -- and this is ignorance on my part. Wouldn't a more patient-centered outcome be whether or not the patient had had a gouty attack during the period of time? Because that's really the goal of treatment is, not targeted at a number, but what the patient's symptoms are.

DR. FITZGERALD: That is harder to get at, and those two values, the uric acid and the recurrence of attacks, are well-correlated.

CHAIR CHOU: Yes, that's an interesting point that Kim brings up. I think it's a little tricky because even with the febuxostat trials, there is still -- it's like it reduced their gout flares by 50 percent or

something. So there's still a -- people are still having flares even if they were being treated. And so, I think that's a little tricky in terms of measuring. For example, rheumatologists might see people who have more severe disease and more refractory to treatment. And so, do you want to measure the process or do you want to measure the outcome, and how can you kind of case mix adjust and those kinds of things? But I think it's an interesting point.

DR. FITZGERALD: Yes, the patient outcomes would be challenging to measure. As you know, a lot of these happen at home. Can be self-reported. So that would be a challenging measure for feasibility/validity reasons.

With tophus, even gouty rheumatologists aren't regularly measuring tophus size to document progression. So again, I think a very difficult measure unless you're in a clinical trial where that's a

specific outcome.

What we have is a highly-correlated intermediate outcome, the uric acid, which predicts tophus and tophus progression and gout attacks. So, if you bring that down, you're more likely that the patient outcomes are going to come down as well.

CHAIR TEMPLETON: And again, I realize that this is easier to measure, however, in one of the previous measures that we discussed this morning one of the definitions of more severe active gout was the number of gouty attacks, and 2 or more in 12 months. So at some level that's got to be a measure because we're using that as a criterion for treatment.

DR. FITZGERALD: That was an inclusion criteria for the measure, which if you miss some people, I think that's okay. As an outcome measure that becomes harder.

CHAIR CHOU: Yes, I mean again, I

think it's -- most quality measures are not predicated on those kind of outcome measures because of this case mix problem. I mean, there are some that are designed like that. But even, for example, like HIV quality measures and diabetes and things like that you can certainly measure outcomes, but without being able to adjust for case mix it's pretty hard to interpret. How do you benchmark a primary care university tertiary care center versus something out in the community? All those kinds of things.

So anyway, at least to my knowledge most of the quality measures are focused in general on process things. There are some exceptions, things like pressure ulcers and stuff like that where it may be more kind of outcome-driven. But anyways --

DR. MATUSZAK: So if I'm understanding the process correctly, what happens when a patient comes in to be treated for gout is that you're initiating --

potentially if they've had severe occurrences and things, that you're starting urate lowering therapy. And then you're actually obtaining these uric acid levels in that acute treatment phase anyways and you're targeting the treatment to some number.

So if you're doing that in the acute phase of disease management, what rationale do you have for doing it once a year or once in 12 months if you've already kind of accomplished that with your initial phase of disease management?

DR. FITZGERALD: So urate lowering therapy is more of a chronic gout management rather than the acute gout management. The acute gout management is pretty much restricted to the NSAIDS and the anti-inflammatories.

DR. MATUSZAK: I'm sorry, I mis-spoke. When you guys were talking about that they're coming in to start urate lowering therapy, you're actively managing people to

bring down their uric acid level. So you're starting at a standard dose, 150 or 300 of allopurinol and then you're titrating up based on what their uric acid level is. So if you're doing this on a biweekly or semi-weekly basis or monthly basis to titrate them to the point where they're under control, what evidence do you have for doing it just once a year? Doesn't it make sense that you would do it to -- treat to a target in the acute -- in the early management of it? I mean, what does it matter 11 months from now if you started uric acid lowering therapy now? Eleven months from now why obtain that value?

DR. FITZGERALD: And so the evidence that I presented earlier in the gaps of care is that patients aren't getting their urate levels monitored and they're not getting their allopurinol or febuxostat titrated.

The goal of this is to try and get people to do that, to try and get to the target. So your goal is to get to 6.8 or

lower, however you do that through the monitoring or titrations, whether it's done however frequently. We set the one year as a reasonable amount of time for someone to follow up with a primary care physician.

If they're in for their gout, maybe they get titrated aggressively like you're describing, which would be fantastic. Maybe they just get put on the 300 that's typically done. But a level gets checked to see if the therapy is effective. And if it's not, then they have a chance to respond. And so they have at least a year to try and get to that rate.

DR. MATUSZAK: In the previous outcome measure that we just addressed previously then you wanted them if you're changing the dose to do it within six months. But if they're on a standard dose, then just to do it once a year. Is that --

DR. FITZGERALD: I'm not -- no, we're not -- so now we're not asking them.

Now we're not telling them how we want them to get there. We're saying you want you to get there.

DR. MATUSZAK: Got you.

DR. FITZGERALD: We had previously suggested that --

DR. MATUSZAK: Thank you.

DR. FITZGERALD: -- if you start a change, you check it and then do that iteratively.

DR. MATUSZAK: Perfect.

DR. FITZGERALD: This is now just saying, okay, get there how you get there, whether you give them 300 off the bat, or preferably you give them 100 or 150 and titrate to the target.

CHAIR CHOU: And do you have any information? I mean, with statins and A1cs, for example, we know that we can't get some patients below whatever target no matter what. You can treat people maximally and they just don't get to the target. Is the same true --

I mean, I know there's that one new drug that we're not talking about here, but with allopurinol and febuxostat are there patients that you really can't get below a target of 6.8 if they start with a uric acid level of 12, that kind of thing?

DR. FITZGERALD: So we talked about this in committee a lot about where the exclusions might be. And again, those are the rare patients more than the exceptions, but patients with renal disease can be more challenging. Dose titration is perhaps not always done ideally. And again, patient adherence, which we've on instruction excluded as part of the design process, are issues.

So there are challenges there. For the majority of patients they really ought to be able to get their uric acid down. The drugs work well. The -- you know, when titrated are effective in lowering urate.

DR. VISCO: Just help me get my head around this a little bit more. If we are

really talking about an intermediate health outcome and not a process measure here, and long-standing high levels of uric acid may lead to chronicity of gouty symptoms, wouldn't it make more sense to measure this in a two-phase way and say -- ask people not only to report the numerator when it's less than 6.8, but also when it's over 6.8, much like you've reported hemoglobin A1c that's been chronically elevated, you know, over 9 or whatever?

DR. FITZGERALD: So we had talked about that and group-level versus patient-level quality measures. And I think the diabetes one that you're referring to, the poorly controlled, that -- I think that's a group level measure. And I think we were directed to me more individual patient measure. And so we focused on trying to treat the target, again because there was rationale in that and there's less -- we would have -- I mean, we could certainly come up with a bad

definition as well. But we focused on trying to get patients to treat to target. It is partly educational.

CHAIR CHOU: If there are no more comments, let me just try to I think summarize kind of where we are at.

So I think like some of the previous measures we've noted that this is -- a lot of this is based on epidemiologic, kind of correlational data without direct evidence. We are looking at an intermediate outcome. Some questions came up about the denominators and the time kind of criterion that are in the measure. I think those were kind of the main issues. And I think maybe we're ready to do a vote on the evidence.

DR. FITZGERALD: Sorry, Roger. I just have the numbers. You were talking about how effective therapy is. And in the febuxostat trials 80 to 90 percent of patients on either of the febuxostat groups had met a serum urate level of six. And that included

patients with CKD up to three. So four was excluded.

CHAIR CHOU: Okay. So we expect people who can't get down to be fairly low because we're using a higher threshold here of 6.8. Less than 10 percent maybe, something like that.

Okay. I think it's time to vote on the evidence. Again, this is a trial measure, so we're voting to whether this should proceed in terms of our evaluation to be tested.

MS. PHILLIPS: Okay. We're voting on measure 2549, the evidence. Your options are one for high, two for moderate, three for low, four for insufficient evidence with exception, and five for insufficient evidence. The voting begins now.

We're at twenty. There. Twenty-one. Great.

All right. Nine for moderate, four for low, four for insufficient evidence

with exception, and four for insufficient evidence.

CHAIR CHOU: So this barely crosses 60 percent.

MS. PHILLIPS: Yes.

CHAIR CHOU: Let's just -- is there anything new that we need to discuss here, Steve, JD?

(No audible response.)

CHAIR CHOU: I think we've --

DR. BROTMAN: -- think so.

CHAIR CHOU: Okay. Any other comments from the rest of the panel?

(No audible response.)

CHAIR CHOU: All right. I think it's time to put this one to a vote as well.

MS. PHILLIPS: All right. We're voting on measure 2549, 1b, the performance gap, with one for high, two for moderate, three for low, four for insufficient. Voting begins now.

We're at 19, so -- there we go.

Twenty. We need one more. You could all vote again. You may have jumped the gun. There we go. Twenty-one. Thank you.

All right. One for high, eleven for moderate, three for low, and six for insufficient.

CHAIR CHOU: That's in our gray zone, I think. Fifty-seven percent. So again, what do we do here?

MS. FRANKLIN: So it's within the 40 to 60 percent range, which is our gray zone. It means that the Committee technically has not reached consensus on this measure. It goes forward though FOR consideration.

CHAIR CHOU: Okay. So I think that's been noted.

The next area is priority. Again, anything to add here, Steve or JD?

DR. BROTMAN: Just the same comment that appeared, previously appeared in this work group, that although gout is a high priority, the work group is not sure that

monitoring to this target represents a high priority gap.

CHAIR CHOU: Any other comments from the rest of the panel?

(No audible response.)

CHAIR CHOU: All right. I think it's time for a vote on priority.

MS. PHILLIPS: Okay. We're going to vote on high priority for 2549. One for high, two for moderate, three for low, four for insufficient. The voting begins now.

We're at 20. So if you can all vote again. We're still at 20.

(Laughter.)

MS. PHILLIPS: Okay. We may have to redo this one. Okay. All right. Well, we were at 20. So -- yes. Okay. This is not a vote, but starting now let's -- everybody vote again and we'll see if we can get to 21.

MS. STREETER: Also when you look at the clicker as you push the button, if you see a red light, I think that means the

battery is either out or going out. So let us know and we can replace it.

MS. PHILLIPS: We're still at 20. There we go. Twenty-one.

Okay. Now we're actually going to vote.

(Laughter.)

MS. PHILLIPS: All right. Now we're going to vote, starting now. This is the real vote.

(Laughter.)

MS. PHILLIPS: All right. We're still at 19. We're still at 20 anyway, so -- is it you, Katie?

Does everybody have a green light on their -- all right. Well, we're still at 20. I'm not sure what's going on with this.

Yes, but let's vote on the actual measure. Yes, maybe we'll get all 21 this time. All right. Now.

Now we're at 20. Twenty-one. Okay. Great.

All right. One for high, nine for moderate, five for low, and six for insufficient.

CHAIR CHOU: Okay. So we're still in the 40 to 60, so this moves forward without consensus.

Then I think we need to move on to the next area. So quality construct isn't relevant, I don't think, for this one.

And we're going to do the specifications a little bit differently. I think basically we just want to -- instead of trying to use the stuff that's on the script, we're going to be focusing on whether we think that the things that are being measured are what should be measured, basically. Is that correct?

MS. PHILLIPS: That is correct. We'll be voting on trial -- on whether the specifications are consistent with the evidence. We won't be of course looking at any of the testing at this time. And this is

not a must pass criterion. It's more just a vote indicating the Committee's feeling about the specifications as presented by the developer.

CHAIR CHOU: So, first before we do the voting, Steve or JD, anything to mention here?

(No audible response.)

CHAIR CHOU: Anything from the rest of any other members of the panel?

(No audible response.)

CHAIR CHOU: I had previously just mentioned some of my thoughts about, you know, potential exclusions. For example, somebody who is stable and doesn't have -- is on uric acid lowering therapy and is having no gout attacks, whether they really need to be in the denominator. And the other one might be people -- considering looking at people with tophaceous gout or erosions separately, or somehow stratifying those populations.

All right. So I think we're

voting on whether we think that as specified it can move forward, right?

MS. PHILLIPS: That's correct.

CHAIR CHOU: Okay.

MS. PHILLIPS: Okay. We are voting on 2549, trial measure specifications. One for high, two for moderate, three for low, four for insufficient. Voting begins now.

We are at 20. We're at 21. Great.

We've got 1 for high, 11 for moderate, 5 for low, and 4 for insufficient.

CHAIR CHOU: So, taken under advisement. That's still in the gray kind of 40 to 60 range.

The next area is feasibility. Any comments about feasibility from Steve, JD or the folks who looked at feasibility?

DR. BROTMAN: Yes, I believe this was the same data feasibility testing summary that we saw before in some detail, unless there's a correction to be made.

DR. FITZGERALD: No, that would be it, but the data summary that you saw before.

CHAIR CHOU: Any questions or comments?

(No audible response.)

CHAIR CHOU: All right. Let's vote on feasibility.

MS. PHILLIPS: All right. Feasibility for 2549. Your options are high one, two for moderate, three for low, or four for insufficient.

We're at 20. There. Twenty-one now. Great. Thank you.

We've got 5 for high, 11 for moderate, 2 for low, and 3 for insufficient.

CHAIR CHOU: Okay. So that passes. It's over our 60 percent. The last criterion is useability and use.

Steve or JD, any comments here?

DR. DANIELS: Just so you know, it's not the after lunch lull. We're just being quiet.

(Laughter.)

DR. DANIELS: Is that it's sort of some of the same stuff on it. It's going to be hard to do this when the finding on what is gout and the levels are kind of all over. So some of the same things. That's where I think they're having trouble.

CHAIR CHOU: Yes, I mean I guess my concern is again about the unintended consequences, whether this is going to lead to unnecessary testing and/or treatment and whether that can be measured, or how that can be measured. So just put that on the table there.

Anything else to add to what's been said already?

DR. FITZGERALD: Just to address that concern that the burden or the cost of a uric acid level wouldn't be prohibitive and most -- the patients would be on urate lowering therapy already, so it might be a dose change.

MS. PHILLIPS: Okay. Voting on measure 2549, useability and use. Your options are one for high, two for moderate, three for low, and four for insufficient information. Voting begins now.

We're still at 20, so if everybody can just make sure. There we go.

Okay. One for high, eleven for moderate, four for low, and five for insufficient.

CHAIR CHOU: All right. This is again in the 40 to 60 percent range. So a process question: So we have several things that fell into this 40 to 60. Does that mean we don't do the final vote, that it just -- we still do the final vote?

Okay. So now we do our vote. Do we want this to move forward as a trial measure?

MS. PHILLIPS: 2549, overall suitability for trial measure. You've got option one for yes and option two for no. And

voting begins now.

Okay. We are at 21.

And 11 for yes and 10 for no.

MS. FRANKLIN: So it's still in the gray zone and we'd still continue forward with consideration of the measure. Again, keep in mind that these deliberations of the Committee would be summarize and report and go out for a public comment and we'll come back and reconsider the measure.

CHAIR CHOU: Okay. So when we reconsider it, are there opportunities for them to revise it, the measure after all the public comments and things, or do we end up voting on the exact same measure again?

DR. PACE: Typically it's the same measure. During that time period of the comment, if the developer has something to offer that you should consider when you have all the other comments and make your final decision on whether to recommend this or not, they can do that during that 30-day comment

period as well. But typically it's the same measure, but there may be some clarifications or potential changes. I mean, the tricky thing is these measures are already specified as e-measures, so it's not something that can change quickly. And so, that's something that we'll have to deal with if they want to make any changes.

DR. VENTURA: Roger?

CHAIR CHOU: Yes?

DR. VENTURA: I guess it was a concern voiced by a couple other people also. I'm not sure that using this algorithm is appropriate for a test measure. Maybe we're being -- the rigor we're expecting is too high for something to be submitted as a test measure.

CHAIR CHOU: Yes, I mean, I don't know if the NQF folks want to address that. I mean, to me it's seems like it is very hard for us to assess some of the latter things that really -- where we're really able to

provide a lot of input before it's been tested is in the evidence kind of performance gap, health priority kind of areas and it's a lot harder for us to really make informed judgments about some of these other things. So I'll see what the NQF folks have.

DR. PACE: Yes, I think -- and we specifically said these measures should meet the importance to measure and report criterion because if they don't meet it now, it's not going to meet it when they come back after testing. And so, that bar should be the same as for any measure that's coming to us for consideration for endorsement.

Regarding the specifications, it has to be an HQMF, which of course it is, and we've already checked that. And I think the big question about specifications is really a connection back to the evidence. So if the evidence the presented; and occasionally we see this, evidence is presented fine, but then the way the measure is specified doesn't

really sync with the evidence. So that's a problem then that you should flag.

And then the feasibility is mostly going to be about can the data be obtained from EHRs and does the measure logic work, which again should be tested. And then useability and use is really, as you've been talking about, it's the potential for use as improvement and accountability.

So the big one is really the importance to measure and the evidence criterion.

CHAIR CHOU: So when you submit a new measure, it's always a trial measure?

MS. FRANKLIN: No, this is only being piloted for e-measures that are in -- without testing, that have not yet been tested.

DR. PACE: And not every e-measure has to go through this. So if an e-measure has already been tested, it just comes in and is assessed against all the criteria.

CHAIR CHOU: All right. I think we're moving onto our next measure, which is 2526, anti-inflammatory prophylaxis with urate lowering therapy.

I think I'll hand the mic to John to give an overview from the developer.

DR. FITZGERALD: Thank you. So the brief description is this is looking at the percentage of patients with gout who are initiated on ULT who are also receiving concomitant anti-inflammatory prophylaxis, be it either low-dose colchicine, NSAID or glucocorticoid to reduce flares.

The numerator would be patients co-prescribed low-dose colchicine, NSAID or glucocorticoid. The denominator 18 years old or greater with established gout initiating urate lowering therapy. Denominator exclusion were patients with contraindications to all co-therapies. That would include the NSAIDs, steroids and colchicine. And those are specified.

CHAIR CHOU: So, Christian and I were the lead discussants on this. So, I'll just give a really brief overview. I'm not going to repeat what John just said.

It's basically as when you're -- because people with -- are at risk for a gout flare when the uric acid level changes, whether it's going down or up. This one is about starting an anti-inflammatory during that initial period.

There were several -- I think the main concern that came up in the work group discussion was that really there was only one trial of anti-inflammatory meds versus placebo. It was small; 43 patients, and it used colchicine. So it didn't address the other drugs that are specified in the measure

There was a much larger trial that was also cited, but it actually compared different durations of anti-inflammatory prophylaxis. It didn't compare prophylaxis versus no prophylaxis. So even though that

was a much bigger study, it didn't directly, I think, address the -- whether to use it. It was more about how long to use it. And that much larger trial was also colchicine, so there was really no evidence presented about NSAIDS and/or glucocorticoids.

There was a denominator exclusion for people with contraindications. So as you all know, NSAIDS and steroids have lots of contraindications, and colchicine has some as well. I think this would be -- I mean, you'd have to have a contraindication to all three, I guess, to be excluded.

I don't know that there was a whole lot of other stuff here. I guess the one thing I would add is at least in the febuxostat trial they did put everybody on anti-inflammatory prophylaxis. I think it was with colchicine. I actually don't remember what they did. But they did something for the first eight weeks. And so at least the trials have been designed to use anti-inflammatory

prophylaxis.

DR. FITZGERALD: Yes, they had used naproxen during that study. So there's evidence of NSAIDs being used, but it's indirect. There's a known risk of a gout flare anytime there's a change in uric acid, so if you go out and have your shrimp and beer dinner, shrimp and lobster, you're going to raise your uric acid and you'll have a gout attack. By the same example, if you were having shrimp and beer every night and you went vegetarian, you would suddenly lower your uric acid and you'd also be at risk for having a gout attack. So a change in uric acid will increase your risk of having an attack.

And so, the baseline risk of -- let's say someone had a 15 to 20 percent risk of having an attack during a six-month period. We'll go up to 30 percent on the initial treatment. So this has been recognized and co-therapy has been recommended, again not just by the ACR, but by other gout mine groups

including EULAR and BSR.

And when studies have been done -- so Sarawate, who we described earlier, looked at reasons for non-adherence. Gout flare after initiating therapy had a twofold risk of non-adherence in patients. So there are risks to ongoing therapy. And so the recommendations have been to use an anti-inflammatory. And we were providing any of the anti-inflammatories that are used as either being colchicine, NSAIDs or steroid.

And it's true that there is only the one small placebo-controlled trial. Given the known risk I think it's very unlikely to do a large trial. The other trial though, however, the short versus long, is on a large study.

Do you remember how many patients? Was it 2,000?

Yes, I think it's 2,000 patient. And so it was three months versus six months and there were flares after discontinuation

after three months. And so the recommendation there was for the six months. So we're not going to be able to get a placebo versus prophylaxis trial. I think the short versus long is the best that we're going to be able to get.

CHAIR CHOU: Christian, do you have anything to comment or add?

DR. DODGE: Just echoing what was brought up in the work group, just that the evidence is mainly for colchicine. And the inclusion of the other options I think makes this a little harder to justify in terms of evidence.

CHAIR CHOU: Comments from the rest of the panel? I see one up over there.

DR. MATUSZAK: Just two quick questions. First one is the most important question I'm going to ask you all day, which is does it appear to be the hops or the barley in the beer --

(Laughter.)

DR. MATUSZAK: -- that seem to be the bigger problem with gout?

And actually the second one; and I'll let you answer the first one, is actually what about the over-the-counter NSAIDs? Is there any way to take those into account? Obviously if people have Aleve, do I have to write them a script for naproxen in order to get credit for this quality measure? Thanks.

DR. FITZGERALD: I think I'll take the first.

(Laughter.)

DR. FITZGERALD: I'll take the Fifth on the first.

Yes, so over-the-counter Aleve twice a day would be sufficient. During the specificity testing we'll see how well we do at capturing that. The EMRs are designed to capture those. If this is specified, people will be noting it more. Again, some of the coding and documentation will be driven by some of these measures.

CHAIR CHOU: Other comments?

MR. SCHUNA: You don't specify duration, but yet one of your studies suggests there's a difference in duration. And I guess I'm wondering why that was. And would one week be sufficient for this prophylaxis?

DR. FITZGERALD: Yes, we had debated about duration. And there was concerns about holding everybody to the six months, and so it was left off. And so, as long as there was some documentation of prophylaxis being done with the initiation, that would be sufficient.

CHAIR CHOU: Other comments or -- and I guess just to follow up on the specific drugs issues, I mean, there really is very little data about use of steroids, is that correct? I mean, not just for this purpose, but just for use -- I mean, there's anecdotal stuff, but in terms of published research stuff there's not very much.

DR. FITZGERALD: Steroids would be

the least published, or at least the NSAIDS we do have evidence of it being used in trials. But again, it was to permit prescribers -- if they would use any of the three anti-inflammatories. In practice it's not used really as a prophylactic agent, and I don't think we would see much of it.

CHAIR CHOU: Well, I mean, I guess I have some concerns about putting a patient on eight weeks of prednisone, for example, for prophylaxis of flares. And I would -- I think it would make me more comfortable at least if it was restricted to colchicine and NSAIDS.

DR. FITZGERALD: Yes, I think we had debated that also because nobody likes the idea of prolonged steroid. And it was left in there for patients who are on existing -- there are a lot of patients who might be one existing steroids for polymyalgia, asthma or other indications.

CHAIR CHOU: Comment?

DR. ANNASWAMY: My question is

more about sort of the specifications, but I guess I'll hold it until then.

CHAIR CHOU: Okay. Other questions about the evidence?

(No audible response.)

CHAIR CHOU: Okay. I think we can move to a vote. So just to summarize, we have the one placebo-controlled trial. It's small, but definitely showed a decrease in acute flares. And then we have that bigger study, which was six months versus eight weeks, I believe, and the six month group had fewer flares. We've already discussed the issues about most of the data really being colchicine with some data on NSAIDs and potential concerns about the corticosteroid component. I think those were the main things.

I guess there was a question about whether there should be a duration-kind of component to the measure, because there currently isn't. That's not within the evidence thing. No, I don't think.

So I think we can move ahead with the vote.

MS. PHILLIPS: Okay. We're voting on measure 2526 for the evidence. One for high, two for moderate, three for low, four for insufficient evidence with exception, and five for insufficient evidence. And voting begins now.

Okay. We're at 21.

MS. PHILLIPS: Eight for low, seven for insufficient evidence with exception, and two for insufficient evidence.

CHAIR CHOU: So, I think we're in the 40 to 60 percent if you add up high, moderate and insufficient with exception.

MS. FRANKLIN: No, we don't count the --

MS. PHILLIPS: High, moderate and insufficient with exception.

CHAIR CHOU: Yes, one, two and four together. So, yes, so we're at 52 percent, so we proceed. So, no, we don't have

consensus yet, but we have enough consensus to keep moving forward.

So the next area is opportunity for improvement. So in the measure work sheet and what was presented to us and in our work group discussion the main evidence presented was this VA study that showed that few patients, 10 percent of whatever, of the VA patients received prophylaxis during acute -- I mean, during the initial treatment phase. I think that was the primary evidence presented, but it was implied that there's a lot of other -- that there's other evidence of similar performance gap there.

Christian, did you have anything to add there?

DR. DODGE: I think the differential between on-demand treatment versus prophylactic treatment wasn't totally clear. What the magnitude of impact was, and just taking these 30 percent increased risk of flares and treating those acutely versus

taking the bulk of people where most of them are being unnecessarily; we just don't know which ones, and burdening them with these --

CHAIR CHOU: Other comments from the rest of the panel?

(No audible response.)

CHAIR CHOU: And, John, did you want to respond to the --

DR. FITZGERALD: I think the main argument for the prophylaxis is the loss of adherence with patients who do have a flare because they then tie -- the data shows that there's a twofold risk of dropping out. And I think what they do is they tie the ULT treatment to their gout getting worse.

DR. DODGE: Understood, but I think that that would be an education piece. When you're starting a therapy you'd want to make sure that they could expect that and then had contingencies for acute management versus -- what strikes me is -- not that it's unreasonable, but I think it strikes me as

heavy-handed to make that a performance measure.

CHAIR CHOU: Other comments?

(No audible response.)

CHAIR CHOU: All right. I think we're ready to vote on the opportunities for improvement performance gap issue.

MS. PHILLIPS: Okay. 2526. You have four options. One for high, two for moderate, three for low, and four for insufficient. And voting begins now.

Okay. We're at 21.

One for high, eight for moderate, nine for low, and three for insufficient.

CHAIR CHOU: All right. We're 40 to 60 percent, so again we haven't reached consensus, but we have enough to keep moving.

So the next area is priority. I think this touches on what Christian said, just how big of a clinical impact is this of avoiding flares but having patients take a drug versus having them treat flares acutely.

And John mentioned that one of the other consequences of having the flares may be that people stop taking their uric acid lowering therapy completely.

And then there's been evidence presented previously about how common gout is and how -- the impact in terms of productivity and health outcomes and all that.

Other comments? Christian?

DR. DODGE: Just the idea that citing costliness of gout flares versus prophylaxis for a much broader group of people, I'm not sure how that would offset the cost of the flares.

CHAIR CHOU: Yes, John, do you want to respond to that?

DR. FITZGERALD: I don't have any data on that. NSAIDs wouldn't be that costly. Colchicine is no longer an inexpensive option, unfortunately. Hopefully will be again soon. And potential side effects I couldn't state. So, gout flares are costly as far as lost

work, visits to the ER. The average cost of a gout patient is usually $300 per year or more than non-gout patients for just their gout-related activities. When you add in all the other sort of comorbidity conditions, then the cost of gout patients go up quite a bit. They can be in the $3,000 or more per patient. So gout flares are costly. It would be hard to say how that one would shake out.

DR. ANNASWAMY: Just to clarify, what is the issue with colchicine?

DR. FITZGERALD: So, colchicine used to be $0.30 a pill, and three years ago the FDA -- and colchicine has been around since Egyptian days originally as an emetic drug, because of that known side effect. In the, I think, it was 1600s it's effect on gout was starting to be used.

In 2009 the FDA branded it for one single manufacturer, who then -- the centuries of research that went into that were then reaped. The price went up to $3 a pill. So

used twice a day, it's now a $6 therapy instead of a $0.60 therapy. And the hope is that once it goes -- it was given a short run of branding, initially three years, but that was extended, I don't know for how long. But colchicine is currently not the cheap option it used to be.

CHAIR CHOU: Other questions or comments?

DR. YAZDANY: Can I just make one comment.

CHAIR CHOU: Sure.

DR. YAZDANY: It's sort of a soap boxy comment. So forgive me, but I do think that as a person who chaired the ACR's Choosing Wisely campaign, sometimes some of the subconscious decisions that we make drive up health care costs. So we want a randomized controlled trial that shows that prophylaxis works. So a company did a randomized controlled trial of colchicine, which rheumatologists have known literally for

decades and decades, and if not centuries, as a medicine that works. And the same thing for NSAIDs.

So, I just think we have to be careful. For a lot of the things that we would have RCTs for drug therapy there has to be a really strong financial incentive for something that's known to be an expensive drug. Some of these things that are very inexpensive and yet very effective based on sort of decades and decades of clinical experience we are throwing out because there isn't an RCT, even though there's very, very strong international and national consensus. So I'm just going to say that. I'm not trying to sway your decision, but I just want to point that out.

CHAIR CHOU: Well, did you want to say something, Karen, or -- okay.

DR. PACE: No, it's been sitting there a long time.

CHAIR CHOU: Yes, I mean, the

point is well taken. I mean, one response would be that. That's what people always say when there isn't evidence. And there are a lot of things in medicine where if we hadn't pushed for the studies we wouldn't have learned that they didn't work or they didn't work as well as we thought they should. And I would argue that a trial of colchicine would be pretty cheap --of colchicine prophylaxis of a couple hundred people would be pretty cheap and not that hard to do. Or a trial with naproxen. But I think some of this is beside the point here.

Let's come back to the voting. So we're voting on the priority. And again, just the high, moderate, low or insufficient here. I think we can move forward.

MS. PHILLIPS: Okay. We're voting on measure 2526, priority. We've already got our options for voting. And voting starts now.

Okay. We're at 21.

We have 2 for high, 5 for moderate, 12 for low, and 2 for insufficient.

CHAIR CHOU: So I think this is just below our 40 percent threshold. Is that right? Sixty-seven percent low, so 33 are -- actually it's less than that. Eighteen and -- ten and twenty-four -- thirty-four percent for high and moderate. So we're a little bit below our 40 percent threshold. So what does this mean? This means that we stop here?

MS. FRANKLIN: Yes, it would not continue forward.

CHAIR CHOU: Okay. All right. So I think we're down with the gout measures.

And now we're moving onto the rheumatoid arthritis measures. And the first one is going to be functional status assessments.

Oh, yes, go ahead. Sorry.

DR. FITZGERALD: Well, I do want to thank all members. It's clear that we could put a lot of time and effort into this.

And thanks to the NQF staff. Throughout the process you've helped get this through.

I do not want this to reflect on my interest on RA. I'll be leaving shortly to catch a plane to the airport.

I just had a procedural question. Was there public comment on the last two?

MS. FRANKLIN: Oh, we don't do public -- well, you explain.

MS. PHILLIPS: Yes, we don't do --

DR. FITZGERALD: Okay.

MS. PHILLIPS: -- public comment. We do it at specific times in the agenda.

DR. FITZGERALD: Oh, okay. With that, thank you.

CHAIR CHOU: All right. Thanks, John, for being in the hot seat --

(Laughter.)

CHAIR CHOU: -- and answering all the questions. I know how that feels.

DR. ANNASWAMY: John, I think you're batting a 500, so you should feel

great.

(Laughter.)

CHAIR CHOU: I heard a request for a bathroom break.

PARTICIPANT: A bio break, a very short one.

CHAIR CHOU: Yes, okay. A bio break. I haven't heard that term before --

(Laughter.

CHAIR CHOU: -- but, yes, I think we can take one.

(Whereupon, the above-entitled matter went off the record at 2:29 p.m. and resumed at 2:36 p.m.)

CHAIR CHOU: All right. So, we're going to try to reconvene here. The next measure we'll be talking about, we're moving away from gout to RA now. It's about Functional Status Assessment, again nominated by ACR.

This is an eMeasure but it's been tested so this is different from the others,

and now we're not voting on whether to test, we're voting on whether to fully endorse. So, I'm going to hand it over to Jinoos to give us a brief overview.

DR. YAZDANY: Great. Thanks, everyone. So, the first measure is 2524, and this is Functional Status Assessment in rheumatoid arthritis. This is a process measure in which the requirement is that functional status is assessed using a standardized assessment and that the result is recorded in an electronic health record.

I just have a few introductory slides. So, rheumatoid arthritis can lead to destruction -- actually, if you can go back just one for me, destruction of the joints, and there's often rapid functional decline. And simple tasks that many of us take for granted, writing, cooking, dressing, walking can become very painful. And we asked the question why is it important to measure functional status? And it all comes down to

one single point, which is that this is the single most important thing to patients. And that capturing Patient Reported Outcomes I think has become a priority nationally. Next slide.

This is a very important study, and it was published in 2011. And it was a very well done analysis that surveyed both patients and rheumatologists, and looked at what things to consider when they're considering changing therapies.

And you'll see that there is some discordance in the things that physicians value and the things that patients value. And you will see that for patients the number one thing is physical function and mobility. For the rheumatologist it is many of the things that we consider under the general concept of Disease Activity, so things like the swollen joint count, the DAS 28 which is a measure of disease activity, and so forth. Next slide.

So, what you may not know is that

rheumatology has really pioneered the use of Patient Reported Outcomes in clinical trials, and in clinical practice over the last 40 years. We've probably been using PROs longer than just about anybody. We have the legacy measures that the Health Assessment Questionnaire, called the HAQ in all of its various revisions, and building on these legacy measures the NIH PROMIS system now has sort of these state-of-the-art versions of these that incorporates some of the old items and adds new items so that we get rid of things like ceiling and floor effects, and we can more carefully tailor the functional status assessments to our patients.

The psychometric evidence base is very strong. It's a method that was very hard to put this type of evidence into the forms which are really more about the process. But, again, there's been decades of psychometric work, probably more in rheumatology than just about anything else showing that these

measures are valid, and they have content and concept validity, that they're responsive to changes, that they're reliable, feasible. And there's been, I think, a lot of work done in health literacy, low literacy populations. Also, they're available in many different languages. And there's also advice by the national/international consensus guidelines that recommend this process of care. Next slide.

Most rheumatologists do not use a formal PRO to assess functional status in practice, although many may do the full in a more cursory way. And measurement infrastructure does enable next steps so we see this as a stepping stone measure. We can't get to outcomes assessment until people are actually measuring this, and we'll talk about why later. We can't do bench marking, we can't go to risk assessment model or do quality improvement without the basic infrastructure of this superhighway of measurement, if you

will. And I think that we can learn a lot from other countries, especially our European colleagues in Sweden, and Denmark, and France, and the UK, and others who really have used this type of measure as the backbone for a National Quality measurement improvement and for a value-based health care design. Next slide.

I decided to clean out because there were lots of issues raised on the testing, but a lot of the workflow challenges were overcome in our testing, and we'll talk about that in more detail. And I want to point out that many EHRs are increasingly including PRO capacity. This is an example, one of the largest EHR developers, Epic, is providing its customers with a PRO application in the 2012 release. There's a library of PROs that are available and you can select or you can add your own PROs, the common short forms and many of the ones that we're discussing are included in that, and the local Epic user books can

program the software for clinical user to define an event and to direct the administration of such PROs. So, this is a rapidly changing field and we recognize this, but there's some challenges to be overcome. But I think that slowly the superhighway is being built, and we'll talk about the validity ratings. Next slide.

I just wanted to mention this because I don't feel it was totally clear in our submission materials that this measure was reviewed and recommended by the Measure Applications Partnership for use in 2015 CMS programs. It's currently in use and programmed into the Rheumatology National Registry which is a chief qualified clinical data registry CMS.

Somebody on one of the teleconferences asked well, if these things are already happening why is NQF endorsement important? And as a professional society, we share the vision that endorsement accelerates

a coherent harmonized measurement status group across the U.S. health care system, so we're motivated to get these things endorsed. So, those are my introductory comments. Thanks.

CHAIR CHOU: So our leads on this one are John Ventura, Kelly Clayton, and Jason Matuszak. Do one of you want to do the overview to start?

DR. VENTURA: Sure, I'll start and then turn it over to Jason and Kelly. I'm not going to repeat the measure. I think it's worth mentioning relative to the rationale they couldn't face the same issue on this one, and that is there's no direct evidence that the process measure itself leads to better outcomes, although there's a lot of indirect evidence relative to the steps in the process such that it's a way to measure responsiveness to treatment, it's a predictor of future disability and mortality, and it provides feedback to both the patient and the provider. And it's associated with an

increased mortality, also, as evidence for the rationale for why this shouldn't be done. When we got to the first -- the comments, it was basically just what I had mentioned, that there isn't direct evidence about the relationship to health outcomes, but indirect evidence for it. There was one question about health literacy, and I think that was addressed by them having multiple tools that had been valid and reliable measures. And anything else you want to add to that, Jason or Kelly?

CHAIR CHOU: Kelly or Jason, do you have additional comments?

DR. MATUSZAK: I guess a couple of the other things that came up was the difficulty of all these patients because a lot of rheumatoid arthritis patients have significant co-morbidities that also are going to affect the functional status assessments. Obviously, we do think they are very important. I think they have to -- I agree

that I think it's the single biggest factor for the patient and for the treatment providers. But I also found it interesting that these treatment guidelines that you guys pulled this from actually didn't recommend specifically doing Functional Status Assessments, it only recommended doing it as part of the global assessment on patients. And then even in the ACR guidelines that they recommended doing it every three to six months in remission or low-activity states, and every one to three months during the active state. So, I was wondering kind of why the discrepancy, why it set a bar at one level, and then describe the quality measure as something entirely different over what the clinical guidelines are recommending. Is it just a set of, you know, artificially low bar even though it's not what your standard of care is, or what you guys determined should be good care.

DR. VENTURA: I'm sorry. I didn't

address the numerator and denominator, the numerator is the percentage of patients 18 and older with a diagnosis of RA for whom a functional status was performed at least once every measurement period, which was 12 months.

CHAIR CHOU: Jinoos, did you want to respond to some of the comments?

DR. YAZDANY: Sure, I'd be happy to. So, let me start with the question about health literacy. This is an area that actually going into testing we were very concerned about because one of our testing sites in particular has a population that draws at least partly from a safety net and there are a large proportion of people that we actually know from other studies that have low health literacy. It's also a population in which are both Spanish and Chinese speakers so there's the added issue of language. So, as we think about handing surveys to patients I think we can't ignore these things.

Now, one interesting glimpse about

what the magnitude of missing data in a setting like that might be, came from the fact that the workflow at this particular site is that every patient fills out something called a Patient Global in which over time they learn that there's a digital analog scale that goes from zero to 100, a happy face, or not happy face -- but anyway, we all have to read to be able to do that, and people just sort of mark how they're doing. So, we have had the experience of 100 percent of patients are able to fill out that Patient Global Assessment, and there was a discrepancy of 6.3 percent between the PROMIS physical function survey and that Patient Global Assessment, so what that means to us is that about 6.3 percent of this population was unable to fill out the PRO. And in those instances there's an additional burden on the staff that have to read the PRO to the patient, or have a family member help them, so I think that's a small number, but it's

something to be considered.

Co-morbidities do affect functional status, and I'm trying to remember what the exact question was about - oh, yes, okay. So, we had many debates over a period of over a year about which functional status assessments to recommend, and they occurred among dozens and dozens of rheumatologists and experts. And in the end we decided that because co-morbidities can affect functional status so profoundly that there needs to be some flexibility in giving clinicians the opportunity to offer different ones. So, for example, you know, if there's a geriatric patient with a severe disability and one who's a geriatric functional status PRO, that's fine. They should get credit for that.

We did recommend four based on expert consensus that would be sort of the default recommendations coming from an archive of rheumatology, but we wouldn't ding people because of co-morbidities that were going to

do something else, so the measure leaves some flexibility.

And then in terms of guidelines and the timing of assessment, if you read through the guidelines, and we know there's 20 different guidelines that recommend functional status assessment, some of them say that this should be offered to patients at least once a year, other ones say that you should do it at every visit. There's disagreement on that, and the reason that we came down at once per year was that nobody said to it less than that, so that is a bar on which everybody can agree. And partly, that has to do with the psychometric properties of these functional status measures. They tend to decline slowly in the chronic phase of the disease, and if something is declining slowly, so for example with the HAQ, if somebody only loses one point or less per year you probably will get that valuable information by just measuring it once a year. Many people do it more often. That

should be applauded. I think that's great, but we're not going to ding people for not doing it every time.

CHAIR CHOU: Other comments or questions from the rest of the panel or from the lead discussants? Yes?

DR. ANNASWAMY: Yes, I think functional assessment as an outcome measure is easy to understand and comprehend, but functional assessment as the intervention is hard to study. So, it is near impossible to overcome the problem of looking for studies where doing a functional assessment has improved outcome. The analogy I can think of is cake. If you are trying to figure out if an outcome of making a cake is good, you would want to eat it, but you really can't assess the outcome of eating without eating it unless you're comparing it to looking at it. And then you're dealing with having your cake and eating it too.

(Laughter.)

DR. YAZDANY: Can I just make a comment about outcomes measures? We - to make another analogy, I think we need to walk before we can run. I'm going to say that for a lot of these measures, so we are in a situation where most people are not even walking, are not even measuring functional status which is the number one most important thing to patients. And although I understand and share the enthusiasm for moving towards outcome measures, we're in a situation where it's hard to even test these measures because people aren't using them, so that's one thing. It's just like building a superhighway.

The other thing that variables for risk adjustment, and we actually went through this exercise so we said if we were going to develop a risk adjusted functional status outcomes measures, and part of this work was actually with CMS because they're very interested in this for meaningful use stage 3. You know, what are the variables that we might

need? Disease duration, fixed deformities, detailed information about co-morbidities. For example, what if someone just had a joint replacement or has another issue? In rheumatology we have a compounding issue of access to drugs. With many of our biologic agents patients have a 30 percent coinsurance so how do you factor in access to drugs on socio economic status? I know the NTHROP has had a big debate about that, which we don't need to talk about today. So, at the conclusion of all of that we decided that we want them motivated to work towards that, and we actually commissioned a paper on Outcomes Measurement in Rheumatology which our colleague, Lisa Sutter, is actively writing right now to sort of set the stage of how we're going to go from here to there, and why we can't do that right this second. But there was universal agreement that this was the right starting point.

CHAIR CHOU: Kelly or Mark have a

comment?

DR. JARRETT: My comment really relates to over 30 years of struggling to take care of tons of rheumatoid patients, and the reality is that you need this measure because it does not coincide necessarily with all our other objective measures, like joint counts. And it is the most important thing. Patients who have to get up four hours earlier to get dressed so that they can button their clothes and go to work, that's what's important to them, not how many warm, swollen joints they have. And it really truly is a measure of how they feel, and it is the most important thing to them. And, yes, it doesn't relate right now directly to outcomes, but I agree 100 percent, if you don't start measuring this now, you get uniformly measured, you can never get to the next step.

CHAIR CHOU: Okay. Kelly, do you have a comment?

MS. CLAYTON: I have to agree with

Mark. Like I said, I come both from the researcher perspective but also the patient perspective, and having done this for almost 15 years, I think the biggest key to filling out some of these functional status assessments and the instruments used is getting the conversation flowing between the patient and the rheumatologist. Yes, these instruments tend to only look at a snapshot in time as them, you know, being unable to do this in the last week versus, you know, the last few months, but it gets the conversation flowing between you and the rheumatologist that hey, you know, over the last three months I've had this decline, or, you know, that leads to a change in therapy, so change in additional prescriptions, things like that, and physical activity changes.

CHAIR CHOU: Thanks. Other comments?

MS. DAVIS: In my department, we work with various outcomes, and in Minnesota

we're actually doing quite a bit of work as well on functional status for surgical procedures. But in all of the work we're doing in Minnesota it is accompanied by qualify of life surveys, as well as functional status because they're different based on the perception of the patient. I'm wondering if that's something you would consider. It's probably not totally relevant to our conversation, but you laid the stage for it, so -

DR. YAZDANY: So, there's overlap of course between physical function and quality of life assessment. In fact, some quality of life assessments include physical function as a single domain. And, you know, I think in a more advanced performance measurement system we would in a much more sophisticated way capture all of the aspects that are important to patients. And, obviously, quality of life is part of that. Within this narrow disease, though, the reason

that we started the functional status is because we have information that that's the single thing that patients care about the most. I agree, that's an important goal, forward thinking, great to hear you're doing that.

CHAIR CHOU: Yes, Zoher.

DR. GHOGAWALA: So, I, too, I want to applaud the effort to sort of systematically collect patient reporting functional status. The numbers that you quoted for the percentage of patients that are unable to complete these forms reliably is very, very low. However, in monitoring of European registries, as well as efforts in the United States where registries incorporate patient reported outcomes type of data the compliance rate is, in fact, low. And one of the things that at least we've seen on spine side is that in academic centers in areas where there's nurses or study coordinators who have helped patients to get these outcome assessments

completed, answer questions and so forth, the compliance rate is much higher. But as we look more broadly at community practice, small private practice and so forth where the time demands on clinicians are enormous, adding this to the workflow may be a challenge. And I'm just curious to hear your perspective, understanding full well that I think that no matter what the hurdles, we must do this. But our challenge is to getting patients to complete these forms reliably, and we probably should think about that.

DR. YAZDANY: Those are all really important comments. I can share our experience in rheumatology, which is that those people who are motivated to measure this have just come up with amazing grassroots innovative ways to do this in clinical practice. I mean, we have people doing everything from literally having like those old-fashioned scantron machines, and people sort of using a number 2 pencil and sort of

feeding it through and getting the outcome, to startup companies that are pushing these things to patients phones, and touch screens, and people getting it through the Epic My Chart, emailed to them. And I think what's really interesting is that if you build the will and the infrastructure, I think we can leave it to people to come up with a solution that will work for them.

We have exact times on each of these measures that are recommended, so in most cases for the patient it's between three and five minutes to fill out the form. And in most cases it's less than one minute to score the form. So, again, I don't want to minimize that. In a 15-minute visit, that one minute might be a challenging thing for a clinician who doesn't have any support staff. But we also know that it's not a huge amount of time either, so at least there's some calibration.

CHAIR CHOU: Other comments?

DR. ANNASWAMY: In your literature

review assessment has PRO been compared to a provider-based assessment?

DR. YAZDANY: So, there's two forms of evidence. One is for randomized controlled trials, and one is from just observational studies. And based on recommendations from Omar Act, functional status assessments are actually now included in a standard way in RCTs in RA. So, we know from those trials that drugs that work move these PROs, so they move in the same direction as objective measures of disease activity, and these composite measures which we'll talk about later. And that data is summarized in some of the psychometric papers that were cited, and that the measures are responsive to change.

There's also observational studies. The largest one is the CORRONA Database. This is a Consortium of Rheumatology Researchers of North America. It's actually based in community practices, and there it's not a randomized controlled trial. There

people just agree to measure. And what we see in that observational study, and I can forward you the citation, is among these 17,000 patients that have been followed for over a decade, we again see that the functional status assessments are moving in the same direction as the more objective disease activity assessments. And that, in fact, just the measurement, which is the only intervention in that observational study seems to lead to no further declines in function, which is really important, and maybe even an improvement. That's a study by Greenberg that was actually relatively recently published in the American Journal of Medicine in late 2013. And I'm sorry it's not included in the materials, partly it's because we started the materials before that paper was public.

DR. ANNASWAMY: So, in these examples you cited there functional assessment reported by the patient compared to provider-based assessment of disease activity. I was

more wondering along the lines of provider-based assessment of function versus patient-reported assessment of function.

DR. YAZDANY: There has been some work done in that area, and as a general rule one thing that's interesting is that if patients are just - and this is based on a study by Jeff Curtis. If patients are just asked about their function without doing a formal survey, so you just say how are you getting along? Which is usually how it goes in clinical practice. I would say that that's the usual care. Physicians will then tend to over-estimate their function, as opposed to having a standard way of collecting it. So, that's the best information that we have in that regard, and I think it makes sense.

CHAIR CHOU: I have a question. You know, the measure is specified as being in counters for RA. And, for example, a lot of patients with RA may have a primary care doctor, but really their RA care is somewhere

else. So, this - you know, it's - I guess what I'm asking, this isn't being targeted at kind of the routine kind of primary care follow-up. It's really people that are being managed, and it's really going to be mainly at rheumatologists the way it's written.

DR. YAZDANY: So, that's a very important point. Again, we went back to both the literature and some qualitative studies. There's a really nice one that Dan Solomon's group did at the break on where they surveyed the primary care community and just asked them whether they are comfortable diagnosing RA, managing RA, starting DMARDs, a whole range of questions along those lines. And I think the results are consistent with our clinical experience, which is that although in the primary care setting many feel comfortable diagnosing them, so I guess the vast majority don't feel comfortable actually managing the DMARDs, so I think for that reason, and the fact that we really wanted to have

accountability measures for rheumatologists, we wanted to limit this to rheumatologists.

CHAIR CHOU: My other question is just how - you know, the numerator specifies using a valid and reliable instrument. I'm just wondering how that's going to be operationalized.

DR. YAZDANY: So, Rachel may be able to speak more to this, but we've applied for LOINC codes for the various recommended measures. There's really four recommended measures. They include the HAQ-2, the PROMIS physical function short form 10, the PROMIS physical function short form 20, and the PROMIS physical function computer-assisted technology testing. Oh, I guess, and there's a fifth one, which is the Rapid 3, which is the MD HAQ. So, those are the ones that we're going to primarily capture. And then I guess there's also an option for other, something like that.

CHAIR CHOU: Right. And I guess my

question, so when it says I use this other measure, how is that judged? Who judges that, and how is that - I mean, there's many, many functional status measures both disease-specific and non-disease specific and, you know, I think that to me poses a challenge in terms of figuring out what meets the criteria or not.

MS. MYSLINSKI: Yes, it is a challenge. The way that we have specified it we're looking for the result of an assessment. So, if it was just, for example, somebody saying - a conversation between a provider and patient, that wouldn't necessarily be captured in that way.

CHAIR CHOU: But it sounds like any measure basically you would accept, so you're not really - the validity and reliability thing won't really be -

(Off microphone comment.)

CHAIR CHOU: All right. Other questions or comments?

DR. VENTURA: I was going to ask the same question. I thought they would come up -

CHAIR CHOU: Yes, the availability -

(Off microphone comment.)

CHAIR CHOU: So, let me just summarize where we're at. I think everyone - I mean, there's no direct evidence about how measuring outcomes impacts patient outcomes, but I think there's a rationale for doing it. But, you know, kind of clinically it makes sense that in order to trap outcomes and all this other stuff - and then a lot of the evidence is, you know, really about reliability, you know, responsiveness to disease state, things like that. That's really where a lot of the evidence, I think, is coming from. It's almost like looking at a diagnostic test where we're looking at diagnostic accuracy rather than how doing the test impacts outcomes.

I'm not sure if we have to use the kind of the algorithm for evidence rules a little bit differently because to me it doesn't really seem to fit the algorithm, unless we drop to the very end.

(Off microphone comment.)

CHAIR CHOU: With exceptions, okay. So, that means we kind of drop to the bottom there.

PARTICIPANT: Box 7, just going on to box 10.

CHAIR CHOU: Yes. So, I think we're probably ready to move to a vote on this measure.

MS. PHILLIPS: Okay. So, you're going to vote on Measure 2524 for the evidence. Your options are one for high, two for moderate, three for low, four for insufficient evidence with exception, and five for insufficient evidence. Giving in just a second, the mouse cursor is being uncooperative. All right. We can start the

voting now.

(Voting.)

CHAIR CHOU: Okay, it passes the evidence. Next is opportunity for improvement. Jinoos already summarized some of that. Anything to add from the lead discussants?

DR. VENTURA: No, other than they did use three test sites to insure variation in implementation and so forth, and 7 percent there was definitely room for improvement.

CHAIR CHOU: Other comments from the rest of the panel? And you can do the vote.

MS. PHILLIPS: Okay. On the performance, one is fine, two moderate, three low, and four insufficient. And voting can begin now.

(Voting.)

CHAIR CHOU: All right. This one passes with flying colors also. These votes are a lot more clear cut than the gout votes. Next area will be priority, so again think

Jinoos already addressed that. I'd see if John or any of the other lead discussants have other comments to add?

DR. VENTURA: No, just to add that it's been added as a CMS top 20 condition.

MS. PHILLIPS: Okay, we're going to vote on priority for 2534, one is high, two moderate, three low, and four insufficient. Voting begins now.

(Voting.)

CHAIR CHOU: Okay. So we keep moving. The next area is the reliability and validity stuff. Essentially, the specifications. Again, I think we're kind of voting, but not really, I guess.

MS. PHILLIPS: Oh, no, this is a full -

CHAIR CHOU: Oh, this is for a full. Okay. So, this is a full measure. Sorry. So, let's step back, sorry. So, this actually has been tested. Maybe, John, if you want to summarize some of the testing information.

DR. VENTURA: Sure. Reliability was assumed to be 100 percent because it's the extraction from EHR. So, they assumed a reliability component, and then used three validity checks. The first was actually using check values or reliability measure between the hand pulled data and the automatic on the EHR pulled data. And it was very, very high correlation between the two.

We also measured face validity by virtue of expert handling census, using what was appropriate, which was considered very high. And then they also did do a - I believe it was like a cap value on diagnosis of RA between the EHR and the hand pulled, and that came out with very little disagreement.

CHAIR CHOU: Anything to add from the other leads?

DR. VENTURA: And according to NQF case validity is sufficient.

CHAIR CHOU: Okay. So, we haven't done this yet, so do we vote on the

reliability first, and then validity? Okay.

DR. PACE: So, in this case, so I know that this gets into our quirky little aspects about criteria, they're right, that we - the Measure Testing Task Force report, because if you're looking at data element reliability and validity, they're very closely connected. So, that's why we accept data element validity to suffice for reliability, as well. So, I would say use those results of the data element validity as your doing your reading for reliability. And because that would be at the data element level, not the performance score, moderate would be the highest score - rating that it could get.

DR. ANNASWAMY: If you go down the algorithm you're at number ten through twelve on the validity level.

DR. PACE: Yes. We're actually going to use the - if you look at the reliability algorithm, I'm just saying we'll use the same for the data element validity as

- you'll use that same information in making your decision about data element reliability.

DR. ANNASWAMY: What would be an example of the reliability -

MS. STREETER: Can you please turn on your mic. Sorry to interrupt.

DR. ANNASWAMY: My bad. What would be an example of a reliability testing that would be at number 2 or 1. We have number 4.

DR. PACE: Okay. So, at the level of the performance score?

DR. ANNASWAMY: Yes.

DR. PACE: Typically, what we see with this is a signal to noise analysis where you actually have the performance scores for a lot of the providers, and you could do a signal to noise analysis. Sometimes we've seen developers do split half reliability analysis, but the idea is that you actually have computed performance scores for enough providers that you can actually do those kinds of analyses. And I don't have it in front of

me, but how many sites did you test this in?

DR. YAZDANY: There were three sites, and multiple providers at each site. And we did not do signal to noise, but instead we look within sites to see whether there was variability between the providers of individual sites, and then also between the sites. And I can tell you that one of the sites that had something like 12 providers had statistically significant differences in performance between the providers at that site, and then others there were no differences between them. I don't know if that's all that you're looking for -

DR. PACE: Yes. No, that's different.

DR. YAZDANY: Okay.

DR. PACE: That's looking at the, you know, distinguishing performance. But, I mean, they're sometimes related. But, anyway, those are the two that we typically see, signal to noise analysis, or split half, you

know, or using some kind of ICC analysis. But, basically, looking at that computed provider level score, not the data, you know, the agreement in the data from two different ways of abstracting.

CHAIR CHOU: Okay. I still don't really get all this stuff, but I think it would helpful at least for me if kind of the NQF people could tell us if they think there are issues with reliability, because I'm not sure that we always know all this random split half stuff, and the signal to noise, and all these other things. I mean, I don't know - I mean, we're often not presented that data, number one. And number two, just how to interpret it I think is going to be difficult for many people. But I'm not hearing that there are any concerns that you guys identified. Is that correct?

DR. YAZDANY: That is correct.

CHAIR CHOU: Okay. And then the lead didn't really identify concerns here

either, so -- and I'm talking about both reliability, as well as the validity component, as Karen said, that these are interrelated.

DR. PACE: And the ceiling is moderate.

CHAIR CHOU: Say that again.

DR. PACE: The ceiling is moderate.

CHAIR CHOU: Okay. Why is the ceiling moderate, again?

DR. PACE: Because the testing is done for the data, not the computed performance score. So, our hierarchy is that that's a foundation but it doesn't tell us exactly about the performance score. And that often happens when you have few sites on which you can actually do the testing.

CHAIR CHOU: All right. So, I believe we're ready to vote on reliability, and then we'll vote on validity.

MS. PHILLIPS: Okay. There are four options. One for high, two for moderate, three

for low, and four for insufficient. We're voting on Measure 2524 for reliability. Voting begins now.

(Voting.)

CHAIR CHOU: We can move to validity now.

MS. PHILLIPS: All right. We are now voting on validity for Measure 2524. Your options are one for high, two for moderate, three for low, and four for insufficient. Voting begins now.

(Voting.)

CHAIR CHOU: So, this passes, so we can move on to the next, which is feasibility. Any comments from the lead discussants?

DR. VENTURA: They did set up especially to look at feasibility, and they were a little bit confused between the two - there were some concerns voiced by the sites. Two of the three said that they didn't think it would feasible from a technical standpoint, although they said it did meet the standard,

and that there is some challenge relative to workflow also. The technical component they thought they could overcome in a matter of weeks, but the workflow they thought might take a few months. They did support it, but there were some challenges.

CHAIR CHOU: Does any ACR folks want to comment on that?

DR. YAZDANY: Just to say that I don't think that we should minimize the fact that this will require workflow changes for individual clinicians. And I feel like I'm very sensitive to that, because I take care of patients, so sometimes it's easy for us to say and hard for people to do.

That being said, it's interesting. When we started this process the subjective feasibility assessments where people just say do you think you can do this, and people sort of typed what they thought were much more negative than the actual implementation of testing. And that was just interesting,

because I think people thought maybe there's a lot of inertia, people are scared of their IT divisions, they don't want them to change the EHR. They think it's going to be hard and expensive, and often I think that those things are true. But on the other hand, many of the testing sites, and actually we continue to see this as we roll out the registry, they are able to overcome many of those things.

And I'll just make a point that, you know, if this measure does get NQF-endorsed and is part of CMS programs, there's going to be a really strong incentive for EHRs to make this easier for docs so that they don't have to reinvent the wheel one at a time, you know, vocally to build this stuff. So, I think that's the vision. I hope it's realized. And I think we have to support, I guess as the ACR part of the job in supporting the members just to make sure that we provide education, tools, and support to make this easier.

CHAIR CHOU: Over here and then down there.

DR. JARRETT: I just have a question for you in the feasibility study that you guys looked at. They did raise concerns about the technical and the workflow changes, but it was interesting because they also - maybe there was a little hesitation on whether or not it accurately differentiated the quality of performance across providers, you know, that there are certain categories that, as you guys did these feasibility things like, you know, they said oh, yes, strong agreement that this will retain its value, and then this one they kind of said ahh, maybe, I guess. Did you get a sense of why the sites felt that it might not be a good differentiator of quality amongst providers?

DR. YAZDANY: It's some of the issues that we've talked about before, which is I think that, you know, this measure is a foundational step, and if we were going to use

the score to differentiate between providers performance without looking at risk adjustment I think, you know, there should be hesitation. But that's not the plan, so I think part of it may be that. Part of it may be that people feel like they're doing this with the are you able to get along kind of question, so they think most rheumatologists on a subjective level think that this part of taking care of patients. You ask them how they're functioning, and maybe they don't believe that doing a standardized assessment is going to be any different than usual care, even though we have some data to show that actually there is a difference. So, I think those may be two issues that are reflected in those qualitative comments.

CHAIR CHOU: Zoher.

DR. GHOGAWALA: So, if I understood correctly, the pilot aspect of this was three sites, in which all three were extracting the data from the electronic health record. The

fact of the matter is that that's not reality in the United States today. So, to look at this from a feasibility standpoint today, to me I think it's very, very worrisome because while it's important to do, while I think that it ultimately must be done, to sort of make this a measure is going to present a hardship for the vast majority of practicing physicians. And an unexpected consequence of this is that the initial enthusiasm and the initial forward momentum on this will be lost. And I'll give you an example.

This is from many, many years ago in Europe with the Spine Tango Registry. The initial pilot study of that showed outstanding patient compliance because it was a group of very, very motivated sites. But as it was brought out to 50 sites with 20 or 30,000 patients, the compliance fell to 30 percent, and the value of that registry is low. So, I think this is something that we need to think about very carefully, because today while Epic

does have a drop-down menu for sure about the common patient-reported outcome measures, it's not in fact installed in the vast majority of sites that have Epic today. And it takes a highly motivated group of people to make it happen, so I don't think this is feasible at least today. That's my view.

CHAIR CHOU: Yes. You know, my only thought is that I have patients who are, you know, any way measured are disabled and are always going to be disabled no matter what I do, and how many times I measure it. And it just raises some questions about, you know, is some of this kind of wasted energy in some patients, at least. And I don't think we need to get into that here but, I mean, I - you know, not necessarily with RA, but patients with, you know, disabilities or other chronic pain conditions and things like that where literally there's nothing - I mean, you know, they've been treated maximally and there's nothing that's going to change their

functional status.

But in terms of kind of the implementation piece, I think my - you know, I already raised the other - the concern I had about what instruments are going to be accepted, and it sounds like you guys are thinking this would be a very broad, basically if somebody uses an instrument that's essentially been published and studied in some way, then that's going to meet the criteria. I think.

Any other comments or questions? All right. I think it's time to make a vote on feasibility.

MS. PHILLIPS: We're voting on feasibility for 2524. One is high, two is moderate, three is low, four is insufficient. And voting begins now.

(Voting.)

CHAIR CHOU: So, we're over 60 percent for high and moderate so we can move on to usability and use. I think some of this

overlaps with the feasibility issues. Are there other comments that the lead discussants wanted to make?

DR. VENTURA: I think just one. There was some overlap with the function measure as a part of PQRS, and the feeling that this is, I guess, pertaining to my question about only using valid and reliable measures.

It was noted that this would add specificity to our PQRSs. I mean, if you open it up to any tool that's possible, how are you adding specificity to the PQRS measure?

CHAIR CHOU: I have a question, actually. I'm a little surprised that this isn't, you know - like for low back pain, for example, we know that people don't measure function and stuff very well in primary care, but in - if you go to specialty clinics they do. That's built into their workflow already, so I'm curious why this is not the case with rheumatology for RA. Is there - you know, it

seems like that would be -- it would be the same kind of situation where there would be a disconnect between primary care management and specialty management, but you'd think that the specialist would kind of be on top of it already.

DR. YAZDANY: I think that there's some clinical inertia on it. I don't know what else to say.

CHAIR CHOU: Other comments or questions here?

DR. ANNASWAMY: One of the unintended consequences that the gentleman over there, I can't see his name, was talking about, but the other is the unintended consequence of setting the bar so low, could it be that, you know, you're finding a lot of people are reaching it, and there may not be an incentive to keep going shooting for a higher bar? That may be one unintended consequence.

DR. YAZDANY: This is more, again,

a vision than the current state of affairs, but the way that the registry is being built is with those nightly uploads of data, and so you are going to be able to sort of create, hopefully, move towards some kind of model that's more learning health care system where you're going to have some people that are measuring it more often, and some people that are measuring it less often, and hopefully get to a point where you can generate data about this. And I know that NQF has, you know, an endorsement and maintenance process, and hopefully, you know, with each iteration of this there's going to be the ability to make this more sophisticated and useful. So, we know this is just a baby step.

DR. DANIELS: Yes, maybe - I'm going to call him Dr. Z, sorry doc. You had some good points and everything but, you know, this kind of tugs at your heart. You kind of want to get this to go through, and I know that just in our setting we sometimes have to

sort of make a quick-draw shot like, you know, this person is presented to you and the family brings him in. Do they go to the nursing home, or they can stay home, they're going to fall and break their hip. And there's a couple of real quick things that we'll do, stand up, sit down, tests that I can do. But to get something that's kind of more complex than that, I guess my question is with -- we're talking about working with groups of people and that. What would be wrong with this, sorry, doc, with just having like an OT or a PT consult. So, if there's a person - I mean, you go and have them see it, and they know a lot more about the function. I mean, you guys are giving people drugs, but as far as functional stuff, they're kind of better than us. So, I mean, what would be wrong with just asking them to do that?

DR. YAZDANY: There is nothing wrong with that. And I think, you know, getting a little bit off topic, but we've

heard a lot about team-based approach to health care, and there are many examples of rheumatologists around the country, and in particular with the care of rheumatoid arthritis where there is a team-based approach. And in that practice is a physical therapist, an occupational therapist, and there's a nursing visit, and some of these things are not done necessarily by the physician, but are done by the co-staff. And I think that in that situation as long as it's in the EHR, there's no reason that the rheumatologist that's affiliated with that practice shouldn't get credit, as well. And, you know, as you move up levels of accountability that practice will look good, and that health system will look good. So, I think there will be incentives, hopefully, you know, depending on how this measure actually is used in the health care system for that to happen.

DR. DANIELS: Do you use a measured

option?

DR. GHOGAWALA: I think that's actually a great idea. The only trouble that I see on this may be a surgeon bias, is oftentimes we see the PT and the OT reports which are very useful for a physical therapist or an occupational therapist to follow someone's care, but we don't really know how to understand it. That sounds silly at one level, but there's no way to quantify the level of disability. And that's something that's very appealing about these validated patient-reported outcome tools. Whichever one you use there are some numbers that have meaning, and you can compare how, you know, one group of patients is doing versus another. And you compare how your patient is doing over time. So, I think there's real value in the patient-reported outcomes approach, as opposed to the PT or OT report, which also has great variability, incidentally, in terms of different therapists and how they fill these

forms out.

But I still continue to - I have real concerns over, you know, how - and I say this as center, we have Epic, so I could see how we could do this, but I have real concerns about programs that don't have an electronic health record that would permit this, so how that would be in the figures.

MS. BURSTIN: Just a quick comment. That's a very fair piece of advice, it's only specified for those who have it, so no one else can be held accountable for using the measure unless you already have EHR. So, again, it's that thinking towards the future, having measures in place when more and more people are on EHRs, but you couldn't do it if you didn't have it. No one will force you to use the measure.

DR. GHOGAWALA: So, this is verified.

MS. BURSTIN: Yes.

DR. GHOGAWALA: Then my concerns

are much less. But there's EHRs and there's EHRs.

MS. BURSTIN: Yes.

DR. GHOGAWALA: So, Epic, and Serner, and so forth have these capabilities, but a lot of people practice with an EHR system that simply doesn't have this capability. So, how would that apply to those?

CHAIR CHOU: I saw some hands up here, and then we'll come back to you.

DR. ANNASWAMY: Another of the unintended consequence, Roger, you talked about earlier which is since, again, this is such a low bar, if a lot of providers started using the other and end up doing a non-validated, non-standardized functional assessment, then we're kind of shooting ourselves in the foot, and we are really not measuring function properly.

DR. YAZDANY: So, thinking again, the onus is on the ACR and the rollout of the registry. And I can tell you that, you know,

they have very dedicated staff that are there to support practices. And I think by putting these tools on line - and, by the way, actually all of the ones that we selected are non-proprietary. Nobody has to pay for using any of these tools, which is really important. For example, the SFM-6 you have to pay for. Right? So, we took all that into consideration, so I think we need to encourage people to use the four recommended tools unless they have a really good reason for using something else.

CHAIR CHOU: And I think that's something that can be measured. Right? We can see how many people are checking off box 4, whatever, and if it's 80 percent, then you know there's a problem, and you need to kind of readdress what we're looking at.

The only other thing I'll mention, and I hope this isn't really applicable here but, you know, this reminds me of when pain was added as an outcome measure to be done as

a fourth vital sign or whatever, the fifth vital sign, which everyone thought was a great idea. And it's probably contributed to the massive, you know, increase in use of opioids. And, you know, so you don't always know what's going to happen with these things, and when you don't have strong evidence linking the intervention with the outcome it really is, I think, important to really follow through and see what's actually happening. Are patients getting better care as a result of the functional assessment, et cetera? There are a couple of other comments here.

DR. MATUSZAK: It certainly sounds like you guys have hit on some usefulness criteria there in terms of some of the overlap with what you're doing with CMS and stuff, which I think is really nice if you can get that type of alignment. I think that's really big, and must be a little concussion piling up on me because I can't remember what the other thing was.

(Laughter.)

DR. MATUSZAK: But, actually, a question for JD. Exactly how far south in southern Illinois are you from?

PARTICIPANT: It depends on the date.

(Laughter.)

CHAIR CHOU: John had a comment.

DR. VENTURA: Yes, I just have experience more with back pain and functional outcome measures, and even two experiences with implementations of the EHR. And while logistically it's a bit problematic, it's definitely the trend, and I think it's going to become logistically much easier. That's been my experience.

CHAIR CHOU: All right. I think we're ready for a vote. Are we ready? Oh, Karen, sorry.

DR. PACE: That's all right. I was just going to say if that's a problem about the specifications and any standardized tool,

I mean, that is an issue. They've got four recommended ones, that's something that the committee can factor into their discussion, or their vote on this measure. But I guess the other thing is, it wasn't clear in the kind of English language specifications, but the measure is specified so that they can't just check that they used, or identify a code for an instrument. They actually have to record a score from that instrument. Is that correct?

DR. YAZDANY: Correct.

DR. PACE: So, that probably needs to be clear in the specifications, because then the point is you will have that data as you've been talking about, so I just wanted to ask basically the committee how they would initiate that.

CHAIR CHOU: At least for me, I think it's acceptable to have the option to use other measures. I just wanted to see follow-up data and if it's a high proportion, then I think it needs to be revisited. If it's

10 or 15 percent, or whatever then, you know, it may not be an issue. So, at least that's my perspective. I wouldn't ask the measure to not go forward or something because of that. Other - yes?

MS. CLAYTON: One last thing to add. One of the articles that was cited in our measure specifically even pointed out that the self-reported measures may be influenced by a patient's mood, self-efficacy, cultural beliefs, so on, but over time you'll see more of a reliable measure versus those random kind of up and down episodes.

CHAIR CHOU: All right. Let's vote on usability and use. And then we'll go from there.

MS. PHILLIPS: We're voting on usability and use for 2524. You've got four options, one high, two moderate, three low, and four for insufficient evidence. The voting begins now.

(Voting.)

CHAIR CHOU: So, this passes and we come to the final vote. And this time we are voting for full endorsement, so not just to test it.

MS. PHILLIPS: All right. We're going to use the additional question slide for that.

(Off microphone comment.)

MS. PHILLIPS: This is for overall suitability for 2524 for endorsement. One is yes, two is no. Voting starts now.

(Voting.)

CHAIR CHOU: All right. We're done with that measure. We have two more. Are we taking a little five or ten minute break? So, maybe we'll reconvene at 5 til, and then finish out the day.

(Whereupon, the proceedings went off the record at 3:43 p.m., and went back on the record at 3:55 p.m.)

MS. STREETER: Thank you. So, as many of you probably know, Ann sent out an

email just announcing that we made dinner reservations for those who are interested in going somewhere after dinner, we thought it would be nice just to have somewhere already booked for you, if you'd like to go. It's two blocks away, and basically we just want a head count now so we can update our reservation. For those who are interested in going to the dinner, if you could raise your hand.

(Off microphone comment.)

MS. STREETER: Okay, thank you.

CHAIR CHOU: All right. So we're going to move into the last couple of measures. I just wanted to apologize because I'm going to have to leave probably after the first measure because I have another phone call I have to be on. But we're going to do Measure 2522 first, so this is rheumatoid arthritis, tuberculosis screening. Again, developed by ACR. The lead discussants are Kim Templeton on the phone, and Linda Davis. And this is for a trial measure, so not fully

endorsing it, just endorsing for further testing. And I'll turn it over to Jinoos for an overview.

DR. YAZDANY: Sure. So, Katie, I'll just have you queue up that slide for me. So, this is tuberculosis testing in people who are newly starting a biologic DMARD.

This is a key patient safety measure, and the rationale is the biologic DMARDs increase the risk of reactivation of latent tuberculosis; that is, tuberculosis that's dormant or asleep and doesn't have any symptoms, and reactivation of tuberculosis is a dreaded event. It can lead to severe morbidity or death, and as someone who took care of somebody who had this two months ago who didn't have screening for tuberculosis before starting a biologic, I can tell you that there was just a tremendous amount of suffering in a patient who almost lost his life. So, this is something that perhaps should be a never event.

Screening detects tuberculosis, and simple treatment regimen generally is effective. Just to give you some perspective because I think some questions were raised about this on the telephone conferences about burden. We think that about, based on CDC estimates, that about 4 percent of the population of the United States has latent TB. It's a much higher percentage among certain racial ethnic groups. So, for example, among African Americans it's thought to be as high as 18 percent, among Hispanics probably 9 percent or higher, and definitely much higher among foreign-born immigrants, as well.

The risk of hepatotoxicity with treatment the usual regimens of either monotherapy with isoniazid or rifampicin in the more recent literature is between about zero to 1 percent, although our experts tell us that with monitoring for liver function tests and early discontinuation if there's a problem, the risk of anything serious

happening is negligible.

Biologics increase the risk of TB reactivation by at least 5 percent, I'm sorry, five times. The signal was initially detected in randomized controlled trials, so this is a serious adverse event happening in the context of over 10 randomized controlled trials. So, I don't think that there's any question that there's an increased risk. Next slide.

We already talked about why this is important. In terms of the evidence, I just wanted to point out that it I think would be unethical to perform a randomized controlled trial of screening. The recommendation is really based on the biology of the disease, and just to tell people a little bit about that.

Our bodies control tuberculosis by forming something called a granuloma. There's a picture of that on the previous slide. And when you give somebody a biologic DMARD, you actually stop that containment mechanism, so

the organism that's asleep is able to wake up and cause all sorts of trouble. There's very strong guideline-based consensus about this.

And then in terms of, you know, people were wondering well, is there evidence that the process of care, so just the screening itself does anything. The closest that I could find is from our Swedish colleagues that have a national registry, and they did a very clever study. And they looked at the background rate of tuberculosis in rheumatoid arthritis patients in two time periods, 2002 to 2006 before there was widespread screening, and then in 2007 and 2011, so that was sort of the background rate.

Biologic starters had a higher rate as expected of tuberculosis, but interestingly over time the biologic users had a decreased risk, so that demonstrates that screening, which is the only intervention. This is not a trial, this is not the regional registry, seemed to protect patients. And they

actually went back and did a chart review in the cases of active tuberculosis that they found, and 16 of the 18 did not have tuberculosis screening documented. So, I think that's probably the best we're going to get in terms of observational evidence.

The gap in care, next slide. The rheumatology registry data which is based on the PQRS measure shows that measuring this has improved performance. We started out at 74 percent performance, and we're up to 93 percent. There were a lot of questions about who's currently participating in the registry, and I just wanted to take a moment and answer that question, because I know that people had questions about that.

So, our current registry has 810 active users. There are 373 unique active sites. There are 31,800 RA encounters, so we estimate that there's about 5,000 rheumatologists in the United States, and I just told you that there's 800 active users.

So, we don't believe that the data from the RCR is necessarily representative of the entire U.S. rheumatology population, and we know that those practices are early adopters, so there's likely more room for improvement from our best guess. Next slide.

I just want to point out in terms of feasibility and validity that this measure is very similar to an NQF-endorsed measure that was put forward by the NCQA in the area of HIV. Their measure reads, "Percentage of patients aged three months and older with a diagnosis of HIV or AIDS from whom there was a documentation that a TB screening test was performed and results interpreted at least once since the diagnosis of HIV infection." So, although the numerator is similar, they actually had a lookback period of forever, which is really hard in performance measurement. But, nevertheless, that's an endorsed measure. We were able to test this measure at one site, and in that one site we

were able to calculate performance. And this had very high expert panel ratings.

Planned uses, next slide. This also has been reviewed and recommended by the Measures Application Partnership for use in 2015 CMS programs. And this is also programmed into our QCDR. That's all.

CHAIR CHOU: Thank you. Kim, would you be able to provide an overview of the measure and the evidence? And then we'll get - have additional comments from Linda.

CHAIR TEMPLETON: Sure. And as was stated, there really isn't evidence directly addressing this from the standpoint of a randomized controlled trial on whether patients are tested or not tested because (a) it would pose a serious ethical issue, and you couldn't do it. And (b) even if you could, the numbers that you would need in a study would be so large that it would be an unwieldy study to do because of the rare incidence of the condition. So, this is something that the

evidence is primarily expert opinion, understanding the risk, as was just mentioned, of putting someone on the DMARD who has latent TB that's not detected and treated initially. So, there isn't direct evidence. However, the consequences of somebody activating latent TB if they're placed on DMARD is so substantial that perhaps we can waive any direct evidence in this area. At least that was part of our discussion when we had our conference call.

CHAIR CHOU: Thanks, Kim.

CHAIR TEMPLETON: Sure.

CHAIR CHOU: Linda, do you have anything to add?

MS. DAVIS: The concern was that the measurement, actually - I mean, this is not unlike all the other measures we've been talking about, the actual process of measurement hasn't been linked to the actual improvement in outcomes, but it has been linked, associated to the implementing the measurement. So, by measuring we improve the

testing of it, which we hope then will be connected to outcomes.

CHAIR CHOU: Thanks. Before taking other comments, I just wanted to note that, you know, this is really a patient safety issue. And it's actually quite rare to have evidence, direct evidence that doing many of the things we do for patient safety like, you know, when we start an anti-arrhythmic checking EKGs and all this other stuff, monitoring LFTs in patients who were put on a potentially hepatotoxic drug, checking their creatinine on somebody who was on metformin. There's no study showing that those things work, but our threshold when it comes to patient safety tends to be different than when it comes to, you know, intervening and doing something. And I would, you know, also point out that the FDA, for example, will pull studies based on case reports and things like that, sometimes very few case reports. So, I do think we look at the evidence a little bit

differently when we're looking at patient safety issues.

Yes, Jason, do you have a comment?

DR. MATUSZAK: Two questions, and I don't remember reading in your guide - or the recommendation, but did you have a preference as to the type of screening method. And did you exclude or take into account in terms of the sensitivity specificities, the percentage, if you know what I mean, like with the Mantelle, for instance, I mean, you're talking about sensitivity sometimes in the .59 range. So, I just wanted to know how you guys decided on what you were going to use, or what you were going to look at, and exclusions for, you know, people with BCGs or stuff like that.

DR. YAZDANY: Those are really good questions. I can tell you that the road to this process measure was long and extremely painful, so just to be perfectly honest.

In our first version of this measure, we really wanted it to be the did you

test, and then did you react appropriately to what that result was? So, it was going to be a more complicated measure, and we really wanted to get there. And we actually wrote the eSpecifications to be able to do that, or tried to write them, and it just failed. So, we couldn't, because of the nuances in terms of different cutoffs, and different labs, and different places that the results are found, and there's just sort of problem after problem in terms of being able to do that second step. So, I think that that's something that we need to work toward, and that's going to require some standardization among laboratories in the U.S., and assays, and just TB screening as a field, in general.

So, therefore, this measure, the only thing that we could actually make work in the current EHR environment is just whether a screening test was performed. And people that have developed workflows to make this work, like for example the testing site, have a

structured place in their EHR where this is recorded. So, I think that's what the workflow is going to end up looking like. Does that answer your question?

DR. MATUSZAK: Kind of. And I was just - the other piece of it that I just had concerns about is just, you know, the prevalence in the population that you're looking at, and then the lack of - I mean, it's a fairly good screening test to begin with, and then, you know, if there is really, truly a horrible outcome that can occur for people, you know, is this actually - does it give us a false sense of security, does it actually accomplish what we want it to, does it actually identify these people when you're looking at these low prevalences, and a sensitivity that's not great. So, is it accomplishing what you're setting out to do, I guess would be a big thing.

DR. YAZDANY: I think those are good points. TB screening is not perfect. You

can have a negative PPD and actually have active disease. You can have a negative QuantiFERON and have active disease. There's a false positive rate, but we had an expert from the CDC participate on the guideline process, and these recommendations follow what the current state-of-the-art is, limited, I agree, and there needs to be more work in this area. But I think the Swedish data are really compelling, and it's showing that there are patient safety gains in an entire country. And I think that that's probably good enough for something that's sort of a devastating outcome for patients.

DR. MATUSZAK: Last piece, and I think that you said in the previous 12 months prior to starting the biologic, why not right at the time you're going to start the biologic?

DR. YAZDANY: So, this has to do with the complicated way that our health care system works. So, let's say that you come in

to see me as a patient and I decide that I'm going to start a biologic on you, and I order a PPD, and you need to have that read 48 hours after you get the test, so that means you might not be able to come back for it until next week. Meanwhile, I'm filling out a prior authorization for the biologic, and that might take months to come back. So, this is based on our qualitative experience, that if it was done in the last year, it's good enough. And you're going to obviously at the clinical encounter see if there are new risk factors or new exposures, but there's just there's all these moving parts, and we didn't want to make it too strict.

DR. ANNASWAMY: What about employee health testing, because this is a very common employee health. Would the availability of that data count falsely towards provider's performance, and if so, if there has been a positive test previously, that is not enough. You've got to now do another test to detect

that latent infection. So, how would you hold the provider accountable for that?

DR. YAZDANY: So, through the consensus process we agreed upon sort of among the rheumatology community that the rheumatologists who prescribes the biologic agent should have documentation of the TB testing or in the case of a person who's not eligible for testing because they've had a previous result, that status documented in their electronic health record.

And, you know, it's imperfect, right, because some people will get it at an employee health center, some people will get it in their primary care office. But, you know, as with all things patient safety we have to have accountability somewhere, and it seems like the person who's prescribing the biologic is a reasonable place for that to land. So, they don't have to do it, they just have to have it in their EHR.

CHAIR TEMPLETON: This is Kim. It's

also noted in there that, you know, if it's in the note it may be based on patient self-report. Is there any data as far as how accurate that is?

DR. YAZDANY: Whether patients accurately report whether or not they've had a PPD test?

CHAIR TEMPLETON: Right.

DR. YAZDANY: I'm sure that that data exists but I don't know it off the top of my head. I can try to look that up.

CHAIR TEMPLETON: Okay, thank you.

DR. JARRETT: Yes. Just a comment again getting back to the utility of the TB test, and just remember that the CDC and most State Departments of Health, if there's a TB exposure in a hospital, that's what we use, that's what we make clinical decisions on, so this is really the state of the science, which is not great, but it is the best we have. And the ramifications of not using the best we have right now are still serious enough that

this has to be done.

And on the self-report side, you know, my personal opinion is much like employee health in hospitals, nobody is going to by self-report, and we don't even let the docs any more read their own PPDs. We make them go to respiratory or someplace to get it done, so I think we have to - we may allow PPDs to be done by somebody else, but there has to be documentation of it.

CHAIR CHOU: Yes, Linda?

MS. DAVIS: I'll reveal my learning curve on RA and DMARDs, but this does not apply to non-biologic DMARDs. Right? Okay. And I trust that that's evidence-based, and everything. Okay.

CHAIR CHOU: Are there other comments or questions? I think that we should be ready to vote on the evidence then.

MS. PHILLIPS: Okay. We're voting on the evidence. One high, two moderate, three low, four insufficient evidence with

exception, and five is insufficient evidence. I'll tell you when you can start. And you can start voting now.

(Voting.)

CHAIR CHOU: I think that's 100 percent that we should move forward. That's a first.

All right. Let's go on to the next criterion. This is performance opportunity for improvement. I think Jinoos presented some of that data. Kim, did you want to comment on that? Did you want to say anything else about this piece?

CHAIR TEMPLETON: No, this again came up during our conference call discussion with the current ACR register reporting. Almost a 93 percent compliance rate, I guess that would be a question for the ACR representatives. How much higher do we think this can go, and will this performance measure continue to drive that higher, or what would they anticipate?

CHAIR CHOU: I thought it was like it went up to 93 percent after they implemented it.

CHAIR TEMPLETON: Right, it went up to 93 percent. Right? But will this change anything by - if we approve this measure?

DR. YAZDANY: So, I think the best guess that we can make just based on the PQRS experience and our registry, so less than one in five U.S. rheumatologists participate in that program, but use of this measure increased performance from the 70s to the 90s. So, that's just - it's just an interesting observation. And what we would anticipate just based on what we know about users of the registry who are early to adopt a lot of these quality improvement initiatives is that performance may be lower in other settings, and that the measurement seems to have driven improvement. So, there's reason to think that there is additional room for improvement, especially if there's more widespread

implementation.

And this is a measure for which, although it's very hard to get to 100 percent, I realize, for really anything, but the goal for this measure actually is 100 percent.

CHAIR CHOU: Linda, do you have anything to add?

MS. DAVIS: I have a question, clarification. The experience that you're describing from the 70s to the 90s was in the registry. Is that accurate? What happened with PQRS data?

DR. YAZDANY: Registries used to report PQRS data. Is that what you mean?

MS. DAVIS: I guess I was under the impression they're two different physician groups that -

DR. YAZDANY: So, people use - people can report PQRS measures outside of the registry, but most people actually use the registry because it makes it easier. Do you know the numbers on that Rachel, like what

percentage of people, rheumatologists use the registry versus not?

MS. MYSLINSKI: We have over 200 providers a year that use it to report.

CHAIR CHOU: Comments or questions from the rest of the panel? All right. We're getting to the end of the day, everyone's tired. I think we can vote on the performance gap issue.

MS. PHILLIPS: Okay. We're voting on performance gap for 2522. Your options are one for high, two for moderate, three for low, four for insufficient. You may begin voting now.

(Voting.)

CHAIR CHOU: Another 100 percent. The next area is health care priority. I think Jinoos already covered this. Kim and Linda, do you have other things you want to add? So, nothing from Linda. Kim?

CHAIR TEMPLETON: No, I don't have anything.

CHAIR CHOU: All right. Other questions or comments from the rest of the panel? All right. I think we can vote on this issue.

MS. PHILLIPS: And you're voting on Measure 2522, priority. One for high, two for moderate, three for low, four for insufficient. Voting starts now.

(Voting.)

CHAIR CHOU: All right, moving right along. So, now we're to the reliability, validity. And I think we go back to how we were doing it for the other trial measures. It's really the specifications, do we think it's measuring the right stuff I think is basically what we're talking about. And I think Jason already brought up an issue, is whether measuring - you know, how people are interpreting the PPD, is the PPD really the correct measure? Do you want to follow-up on that, or have other thoughts about that?

DR. MATUSZAK: I'd let the lead

discussants answer first, and then I'll -

CHAIR CHOU: Okay. Linda and Kim, do you have additional comments here? Nothing from Linda. Kim?

CHAIR TEMPLETON: Nothing, no.

CHAIR CHOU: Okay.

DR. MATUSZAK: I think that the - well, I guess the - let me think for a second. Hold on a second.

CHAIR CHOU: It was just a follow-up on what you had brought up earlier about whether measuring PPD is really what we want to be doing, are there other measures that would be better for screening people, I think is what you were bringing up before; that there's difficulty in interpreting PPDs, that the test itself is not perfect, et cetera.

DR. MATUSZAK: Yes. No, I think that summarizes what I was saying before. There's a couple of different methods, obviously, for screening for TB. And, you know, we don't say any one particular method,

we don't even talk about whether it's a two-stage PPD, or a one-stage PPD, or anything else like that. We kind of leave it up to the individual clinicians to make a judgment on that, you know, so that's part of it.

The other piece of it, too, is the -- you know, even I think some of your sites mentioned this, again, in your feasibility. I guess we haven't gotten to feasibility yet, but is it really accurately portraying the quality differences between the providers? You know, they might be getting a number of patients that have already been pre-screened for other reasons or something else, because whatever it was, like most of your sites actually indicated that they felt like it actually lost validity when it was translated into a quality measure, and that it did not accurately reflect quality differences between providers.

DR. YAZDANY: So, I think that those comments reflect sites that have not yet

worked out the workflow to capture this information reliably. We talked about that a little bit before, so if you just take an EHR that's not trying to sort of move the patient safety needle on this measure, there may be instances where a PPD result is scanned as a PDF, there may instances where the QuantiFERON was done at an outside lab, but there may be - so I think that, you know, as currently structured if you were just going to, you know, apply this and the gold standard was, was the test done anywhere in the health care system, then it's true that it would not reflect accurately the quality of care that was being provided.

Now, that being said, I think that our testing site that has established workflow to get this done demonstrates that it actually can be done, and that there is a slight performance gap. I think that they were in the 80s. This is a research interest of mine, so I can share with you additional data since

this submission that we did at our VA Hospital in San Francisco, and that we did - you know, this was a quality improvement project, 138 patients who had started a biologic in the last year. Of those, 27 percent had not gotten a PPD that was findable through looking for either a PPD that was administered or a QuantiFERON test. And of those, 13 percent were really not completed, and 14 percent was presumed completed but the biologic was started outside the VA, so we think that it might have been done at some other site. So, we had a lot of discussions among the clinicians, if the person at the VA is prescribing the biologic, even if they're renewing it, should the documentation of a PPD reside at that facility, or is it good enough to just continue a medication someone else started somewhere else. And I think that, you know, for patient safety you have to avoid the Swiss cheese, and we decided that a redundant system in which the prescriber of a biologic

has documentation within their EHR would be the safest strategy for patients. So, there are feasibility issues. They want to acknowledge them and not minimize them. But I think that the vision in terms of making health care safer is to make the prescribing doctor accountable by building in the workflow to put it into the EHR.

DR. MATUSZAK: But does just the having to have a test documented in the EMR actually make the provider more accountable? I mean, what's - there's no follow-up afterwards to insure that it wasn't a false negative test, and that your latent TB didn't activate, presumably. So, how do you translate this into actually meaning that the providers are providing better care?

DR. YAZDANY: That gets back to the original comment where, you know, I think in the future the goal is to do the measure that's - we found it, and then this is what we did about it. But we just couldn't actually

make that work in the testing, so I understand the desire for that. We shared it, but we just couldn't make it work yet. So, I think as more people implement this maybe we'll actually through the registry or whatever be able to figure out an innovative way to do that.

CHAIR CHOU: Mark?

DR. JARRETT: I see your point about attribution of the quality factor to the provider, the rheumatologist, but I think it gets to the concept that we need to think not in terms of attribution, that you actually do something, but that you've reviewed the data, had the data, and that's the whole purpose of meaningful use and transfer information is the fact that if it does exchange, and the fact that you documented it and hopefully paid attention to it if it was positive, means that that is a quality factor, because up until now documentation in the charts of what went on any place else other than your own office, let alone in your own office, has been poor to

less than poor. So, I think really at this stage for American health care, I think that that actually is an attribute of quality.

CHAIR CHOU: Other comments about reliability and validity, or again about the test specifications? So, just following up again on one of Jason's points. So, I mean, would there -- it seems like to me it would be fairly easy to require a two-stage PPD or some - you know, you could add some stuff into the measure fairly easily, it seems to me, but you guys just decided you can't go there? Okay.

DR. YAZDANY: We actually really did try. We can share those results with the committee if they are interested.

CHAIR CHOU: Other comments? All right. I think we're ready to vote on the reliability and validity piece.

MS. PHILLIPS: Okay. We're voting on 2522.

CHAIR CHOU: Specifications, sorry.

MS. PHILLIPS: Trial measure specifications, or the specifications consistent with the evidence. One is high, two moderate, three low, and four insufficient. And you may begin voting now.

(Voting.)

CHAIR CHOU: So, this passes. The next area is feasibility. Kind of touched on. Kim and Linda, do you have other comments you want to make about this? Nothing from Linda. Kim, other additional comments on feasibility?

CHAIR TEMPLETON: There was a discussion, I know, what happens as we talked if this test is done elsewhere, but it was thought that that could be accommodated with the EHR.

CHAIR CHOU: Great. Other comments from the group? I'll just say one thing, and that's that I'm not convinced that people read the induration correctly, that they'll read the redness and stuff like that, so I do think that the likelihood of false positives is

there, but that's a problem with TB testing in general, not specific to this measure.

All right. Do we want to go ahead and vote on the feasibility then, if nobody has other things - anything else to bring up? All right, let's do it.

MS. PHILLIPS: Okay. We're voting on feasibility for 2522. One is high, two moderate, three low, four insufficient. And voting begins now.

(Voting.)

CHAIR CHOU: That passes the feasibility test, and now we're to usability and use. Jinoos already presented some data about, you know, its implementation in the survey she cited, and some evidence about improvement in testing rates. And I think Jason and I have both alluded to unintended consequences. Are there other comments from Linda or Kim?

CHAIR TEMPLETON: One of the comments that came up in discussion is if, you

know, someone has a negative TB test and is started on a biologic, and needs to switch to a different biologic, does the TB test need to be repeated, or is the initial test adequate?

CHAIR CHOU: Jinoos.

DR. YAZDANY: So, the eSpecs would indicate that the initial TB test is - there's a measurement year, and if a biologic is discovered in the measurement year then we look back one year from the start of that initial biologic.

CHAIR TEMPLETON: Okay.

DR. YAZDANY: It would not - just to be clear, they don't need another test if they switch their biologic.

CHAIR TEMPLETON: Okay. All right, thank you. That's it for me.

CHAIR CHOU: So, I was just thinking why wouldn't you - so, if somebody's been on a biologic for five years, why wouldn't they - I mean, is there no - I mean, can't you get TB? I mean, it's just like

a health care worker has to be tested every year, or whatever.

DR. YAZDANY: So, at some point, you know, another quality measure might be that you document that you've assessed new TB risk factors every year.

CHAIR CHOU: Other comments? Yes?

DR. ANNASWAMY: This is where I think, again, unintended consequences could be an issue. For example, the false positive test results could be an issue, false negative test results could be an unintended consequence. This is also a safety of the safety measure. TB testing has adverse reactions. In a setting of a previously unknown exposure, a TB test can cause skin sloughing, and so - and QuantiFERON is preferable, so you might end up causing harm trying to minimize harm. So, unintended consequences can be a problem with this.

CHAIR CHOU: Jason?

DR. MATUSZAK: And I think it was

my point, too, is that I think that you guys actually put in here that there were no risks of unintended - or no risks associated with this, and I think that there could be. I mean, you're talking about, you know, positive test leads to six months of INH, liver problems, and those types of things when it might not even have been a true positive test. And I think that you do have to consider that.

But just incidently, just so I know, if they've been treated with six months of INH, can they then start on some biologic, can they do it at the same time? Do they have to wait? What's the -

DR. YAZDANY: That's great. So, I want to apologize for not filling out the unintended consequences section appropriately enough. In retrospect after everybody pointed it out, I think there is actually more literature that could have been cited there, including the exact incidence of side effects with TB therapies and all of that, so I

apologize for that. That data is available, and I've cited some of it for you.

In terms of safety of starting biologics after TB treatment has commenced, the ACR guidelines, the 2012 Sing, et al guidelines say that any time after one month of completing TB therapy consensus is that you can start. Some people wait longer, few people wait shorter, but that's the guideline.

CHAIR CHOU: If there are no other comments, I think we can vote on the usability and use issue.

MS. PHILLIPS: Okay. We're voting on usability and use for 2522. One is high, two moderate, three low, and four insufficient information. Voting begins now.

(Voting.)

CHAIR CHOU: So, this passes, and I think we're to the final vote. So, this is whether to recommend as suitable for a trial. So, again, not for full endorsement, but for further testing. Any final comments? All

right, let's do the vote.

MS. PHILLIPS: All right. Does the measure meet NQF criteria for a trial measure? One yes, two no. Voting starts now.

(Voting.)

CHAIR CHOU: All right, this one passes. So, we've got one final measure. I have to step out, so I'm sorry about that, but I will see you all in the morning.

MS. FRANKLIN: Thank you, Dr. Chou. So, that moves us on to our next measure, and if we can have a brief overview of the measure. This is Measure 2523, Rheumatoid Arthritis.

DR. YAZDANY: Great, thank you. And, Katie, I'll just you queue that up for me again. So, this is a measure requiring that patients, adult patients with rheumatoid arthritis have a standardized disease activity measurement greater than 50 percent of the time in their outpatient RA encounters. So, I think it is worth it to spend just a minute

talking about what we mean by disease activity, because I do think that this is a somewhat complex concept.

This is key outcome in rheumatoid arthritis. It's the cornerstone for advancing therapy. It's the outcome that's used in all randomized controlled trials. It's what we use either using a standardized form or subjectively to advance treatment in clinical practice and judge the effectiveness of our medications. And there's strong consensus that this is something that should be measured in national and international guidelines. Next slide.

So, what is RA disease activity? Well, what's interesting is that because there's no biomarker for rheumatoid arthritis, we don't have a blood pressure or hemoglobin A1C. What has evolved over many decades are these composites of different values. So, for example, it might include swollen and tender joints, inflammation on laboratories,

declining or poor function. And just to give everybody a visual representation of this concept of RA disease activity - oh, I'm sorry, if you could advance the slide one more for me. You'll see that there are six ACR-endorsed measures, and each contains different components. All six include a patient-reported outcome. Some of them also include laboratory features, some of them also include a provider's global assessment. And I know that this is somewhat confusing, and I'll explain a little bit more about this in just a second. Next slide.

So, it has taken us a very long time, and it's been an arduous process to actually just get rheumatologists to agree on what the endorsed measures for RA outcomes should be in terms of the ACR's endorsements. And in this document, which I cite in the materials by Jackie Anderson, citing the ACR's endorsement process, there is a three-year national effort to define and agree upon valid

disease activity measures in rheumatoid arthritis.

And to just give you a sense of the scope of this work, there were 63 existing measures, and we took those through psychometric evaluations, and did surveys of rheumatologists, and saw which ones people were actually using, and which ones were feasible for use in clinical practice, which ones actually had cut points for remission, low, moderate, and high disease activity so we could do some comparisons across the measures. So, that's how we arrived at the six measures. So, even though they include different components, it was felt that they captured this concept of disease activity using these different data elements. Next slide.

There has not been a randomized controlled trial comparing measuring disease activity to not measuring disease activity. What we have is randomized controlled trials that look at a treat to target strategy, and

it's been very interesting because over the last decade, trial after trial demonstrating that it doesn't even matter how you reach that target, but as long as you aim for a target it seems to be that patients do better. And there's new data from the CORRONA Registry, again, this is this observational study with 17,000 patients in which really the only intervention was measurement, because people actually have to submit the disease activity measures. And we see this really nice improvement in disease activity scores over time.

This table shows that this is a disparities sort of sensitive measure, and that disease activity is higher both at the start of the study and later in the study in racial and ethnic minorities; although, over time there was improvement in all groups with measuring, even though the disparity persisted just in terms of the absolute levels. Next slide.

A majority of rheumatologists don't routinely measure disease activity in clinical practice. And in the most recent study at Brigham, you know, one of our academic institutions we found that only 29 percent of the time was one of these standardized measures actually used in clinical practice. Next slide.

This is another thing where although there was concerns about implementation of workflow, all of the testing sites were able to implement this over a period of weeks. And I'll just point out that there's a range of options, so some of these measures require that there's laboratories available at the point of care, not everybody is going to be able to implement that. Some of these have only a patient-reported component, so for sites where the workflow implementation might be a challenge, that's an option, so leaving that flexibility was really important.

Next slide, the planned uses. This

measure has been reviewed and recommended by MAP for use in 2015 CMS programs. It's also programmed into the registry. And this is the one measure that CMS is actually working on to include in meaningful use Stage 3. They've contracted with Mathmatica Policy Research to develop this measure. Mathmatica has been consulting with the ACR, and they've agreed that these six disease activity measures should be the base of this measure as we work towards a strategy of outcomes measurement. Although, I think that they agree that it's hard to do that right. And I'll stop right there.

MS. FRANKLIN: Okay, thank you. So, our lead discussants for this measure are Puja Khanna and secondary is Marcie Harris Hayes. And, also, I wanted to check and see if I have Kim Templeton still on the line?

CHAIR TEMPLETON: Yes, I am.

MS. FRANKLIN: And, I also wanted - if you wanted to help facilitate this,

please feel free to do so. I wanted to give you that opportunity.

CHAIR TEMPLETON: Thank you. I appreciate that.

MS. FRANKLIN: Other than calling on people, which I can do.

(Laughter.)

MS. FRANKLIN: Thanks, Kim.

CHAIR TEMPLETON: Sure.

DR. HAYES: So, actually I'm going to take the ball on this one, and Puja is going to step in to save me when needed.

So, Jinoos did a really nice job of basically doing what I'm supposed to do next, but I'll just try to point out a few things.

So, this is a process measure at the clinician level. And, again, the description of the measure you have up on the slide. I think she represented it nicely. As far as the evidence is concerned, she also presented that, as well. And from the Work

Group meeting we felt that the evidence to support the actual process itself is not available, but the evidence to support tight control of disease activity was strong and appropriate, we felt. Dr. Khanna, do you want to add anything?

DR. KHANNA: One of the things that was brought up, Jinoos, was one of the members thought why was the cut off of 50 percent of the patients used? So, if you can elaborate a little bit on that, I think that would be helpful to the group.

DR. YAZDANY: So, although it seems somewhat arbitrary, we based that on a couple of things. The first thing, and perhaps the most important factor is that we have a range of adoption of this in the U.S. health care system right now. And there are facilities, Geisinger is a wonderful example, that have implemented this workflow over the period of a number of years, so they really had time to sort of test this and improve. And it's

interesting to see that performance at those sites is not 100 percent, nor are they aiming for it to be 100 percent. And the reason is that there may be instances where a rheumatoid arthritis patient is coming in follow-up where they just need an injection, or they have an acute issue, and you don't need to do the whole disease activity assessment. So, this provides room for those instances where people don't need to have a full evaluation, which comes up not infrequently, and sort of empirically based on evidence ruling this out.

The other piece of this is that I think health care is messy, so asking for a complicated workflow to be done at every single visit when, you know, the patient may not want to do it, and the MA may not be able to give them their patient-reported outcome to fill out, that's a really, really high bar. But if you're at 50 percent, it means that you're doing it most of the time. And we felt like that was - it's actually really pushing

the field forward, but not being unreasonable. So, anyway, that was the justification for the greater than 50 percent.

CHAIR TEMPLETON: Any other questions or comments?

MS. FRANKLIN: Marcie had a question, and then - Thiru?

DR. ANNASWAMY: The measure worksheet summary of all the Work Group calls, the assessment of evidence is all over the place. What is the summary of the assessment of evidence by the lead discussants?

MS. FRANKLIN: Marcie.

DR. HAYES: So, I can address that. So, when we went through the work - as we were going through the process up to our first meeting we weren't really certain about how to judge the evidence when it wasn't - the evidence wasn't directly measuring the process. So, it's my understanding that those individuals that rated the evidence low, they were basing it on there isn't evidence,

evidence hasn't been given on the specific process measuring this to all does it affect outcomes? That's my understanding. The Work Group can jump in and -

DR. KHANNA: That is correct. So, if you want a straight answer, the evidence supporting this measure is strong. But it's just that we needed a little bit more clarification on how the evidence was derived.

CHAIR TEMPLETON: Does that answer your question?

DR. ANNASWAMY: Yes, Kim. Thanks.

CHAIR TEMPLETON: All right, thanks. Any other questions or comments?

MS. FRANKLIN: I don't see any around the table either.

CHAIR TEMPLETON: So, are you ready to vote on the evidence?

MS. PHILLIPS: So, we're voting on the evidence for Measurement 2523, one high, two moderate, three low, four insufficient evidence with exception, and five insufficient

evidence. And hang on just one second. Okay, voting begins now.

(Voting.)

CHAIR TEMPLETON: So, it sounds like we have enough to proceed.

MS. FRANKLIN: That's correct.

CHAIR TEMPLETON: All right, good. Now, I'll turn back over to lead discussants on gaps in care and opportunities for improvement.

DR. HAYES: So, the developers provided some information on performance scores that they measured three different entities, and there was variable performance in meeting this criteria that ranged from 35 to 61 percent. They also reported on some data through the ACR's Clinical Registry to show that there was some improvement from CY 211, 43 percent up to 54 percent in 2012, so it appears there is a gap and that it could be improved.

CHAIR TEMPLETON: Okay, great. Any

other comments or questions? Okay. Sounds like we're ready to vote on gaps in care at this point, opportunities for improvement.

MS. PHILLIPS: Okay. Measure 2523, performance gap, one high, two moderate, three low, four insufficient. And the voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay. Moving on to priority, and turn it back over to the lead discussants on priority.

DR. HAYES: For priority it looks like the Work Group rated it at moderate or above. National Priorities Partnership, as mentioned before, has listed RA is one of the top 20 high-impact conditions.

CHAIR TEMPLETON: Okay. Any further comments or discussion regarding priority? Okay. If not, we can go ahead and vote.

MS. PHILLIPS: Measure 2523, priority. One is high, two moderate, three low, four insufficient. Voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, wonderful. Move on to reliability and validity. Turn this back over to the lead discussants.

DR. HAYES: So, we are - this is for endorsement. Right? So, we will go through these steps. So, for reliability, again, this is an eMeasure, so we are to look to validity to make our assessment about reliability. Is that true?

DR. PACE: Right. They did validity of the data elements which is acceptable.

DR. HAYES: Okay.

DR. PACE: And we'll use those results for both the reliability and validity. They did the agreement between the eMeasure data and instruction.

DR. HAYES: So, validity testing similar to what was reported previously, they tested it in three different ways. Performance measures for validity where they abstracted data from randomly selected records. The

second was critical data element validity. This is the part where I struggled with a little bit because the validity I'm used to is very different than this validity, so help walk me through this.

And then, also, systematic assessment of face validity, which there was strong face validity based on expert opinion. And then - and part of the validity measure that did demonstrate it that sites actually did perform differently with the measure; however, there were some questions raised about how specifically you can demonstrate how each site performed differently.

CHAIR TEMPLETON: Any additional comments or questions, discussion?

DR. PACE: So, I'll just answer your question about why this is at least thought of as validity, is that for some of these measures if you're looking at the data, the question is are the data the correct data? So, typically, we've seen this with claims

data. So, the question is when you look at claims data, is that - if you would compare that to "the authoritative source," which would be the medical record, would you get the same result? So, we typically have seen this kind of analysis with claims data often reported as sensitivity and specificity rather than CAPAs. So, that has - that kind of concept has been extrapolated in this eMeasure environment where you're looking at eMeasure -- data that's pulled using the eMeasure specifications, and comparing it to data that would be abstracted from the entire medical record. So, in one place you've got a computer program pulling the data and computing the performance measure, and you're comparing that to some human actually going through the whole record to looking for the data to see if they're convergent. And that's, you know, maybe not totally satisfying, but it's kind of where we're at right now with eMeasure testing. So, if anyone has better ideas, I'm

sure all the measure developers would love to hear it and so would NQF. So, that's kind of where things are in terms - so, it's really the concept of that data being used in the measure compared to the authoritative source, which in this case would be really having access to the full record, not just what the computer program is dealing with.

And just to remind - same as our last conversation, because these are tested at the data element level according to our algorithm the highest rating is eligible, four is moderate. And that would be for both the reliability and validity, because even though they've done face validity, we don't bump it up because of doing face validity.

CHAIR TEMPLETON: Any additional comments or discussion?

DR. KHANNA: So, if I may, this is Puja here. The CAPA was reported as .81 so that's pretty decent.

CHAIR TEMPLETON: So, we're ready

to vote on reliability?

MS. PHILLIPS: Voting on reliability for 2523. One high, two moderate, three low, and four insufficient. Voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, thank you. Discussion on validity, any further discussion on that?

MS. FRANKLIN: None here in the room.

CHAIR TEMPLETON: Okay. Let's go ahead and vote on validity.

MS. PHILLIPS: Voting on validity for 2523. One high, two moderate, three low, and four insufficient. And voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, great. Thank you. Now, turn the discussion back over to the lead discussants on feasibility.

MS. FRANKLIN: So, for feasibility I think we've already kind of talked about

some of the feasibility issues here, so Zo mentioned some earlier. So, those are the only comments that came up with the Work Group.

CHAIR TEMPLETON: Okay. Any further comments or discussion regarding feasibility? Okay, we can go ahead and vote on feasibility.

MS. PHILLIPS: Feasibility for 2523. One high, two moderate, three low, four insufficient. Voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, great. Thank you. And then usability and use, again we'll turn back over to the lead discussants.

MS. FRANKLIN: For usability and use, some of our previous comments have been addressed already during this discussion. And, in general, the Work Group had no major concerns.

CHAIR TEMPLETON: Okay, great.

MS. FRANKLIN: Other than what we've already discussed.

CHAIR TEMPLETON: Any other

comments or discussion regarding usability and use? Okay, let's go ahead and vote on that.

MS. PHILLIPS: Okay. We're voting on usability and use for 2523. Your options are one high, two moderate, three low, and four insufficient information. Voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, great. Thank you. Any other comments or discussions on 2523? If not, I guess we'll proceed to voting on the measure overall.

MS. FRANKLIN: There's none in the room.

CHAIR TEMPLETON: Okay. All right. Let's go ahead and proceed with the vote then.

MS. PHILLIPS: Okay. We're voting for overall suitability for endorsement for 2523. Your options are one for yes, and two for no. And voting begins now.

(Voting.)

CHAIR TEMPLETON: Okay, great.

Thank you.

MS. FRANKLIN: So the measurers passes and that concludes our measure review process for today. But at this time, we want to open the line for public comments or comments in the room.

OPERATOR: At this time if you would like to ask a question please press *1 on your telephone keypad. There are no further comments at this time.

MS. FRANKLIN: All right. So, hearing none I'd like to thank the Committee members and the developers for a very productive day, and we look forward to seeing you tomorrow morning at 8:30 for breakfast, 9 a.m. we'll start with the measure evaluation again. Tomorrow we'll also be addressing harmonization issues, as well as going around the table to discuss gaps scenarios that we'd like to see future measure development.

I'd also like to thank you, Kim, our Co-Chair for being on the line and hanging

in there all day.

CHAIR TEMPLETON: Thank you. I appreciate it.

MS. FRANKLIN: And thanks to Dr. Chou in absentia. With that, I'll give you some time back. We finished early.

DR. DANIELS: Should we leave our clickers and our name tags just here?

MS. FRANKLIN: Yes, please leave everything here except your computers.

(Whereupon, the proceedings went off the record at 5:04 p.m.)

D

J

M

S

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Musculoskeletal Measures Standing Committee

Before: NQF

Date: 05-07-2014

Place: Washington, D.C.

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter