National Quality Forum
Patient Safety Measure Evaluation Web Meeting
Wednesday, February 16, 2022

The Committee met via Video Teleconference, at 10:00 a.m. EDT, John James and Donald Yealy, Co-Chairs, presiding.

Present:

John James, PhD, Co-Chair
Donald Yealy, MD, FACEP, Co-Chair
Emily Aaronson, MD, Massachusetts General Hospital
Joel Bundy, MD, FACP, FASN, CPE, Sentara Healthcare
Elissa Charbonneau, DO, MS, Encompass Health Corporation
Curtis Collins, PharmD, MS, St. Joseph Mercy Health System
Theresa Edelstein, MPH, LNHA, New Jersey Hospital Association
Terry Fairbanks, MD, MS, FACEP, MedStar Health
Jason Falvey, DPT, PhD, University of Maryland School of Medicine
Sara Hawkins, PhD, RN, CPPS, Eastern Idaho Regional Medical Center
Bret Jackson, The Economic Alliance for Michigan
Laura Kinney, MA, BSN, RN, CPHQ, CPHRM, CPMA, CPC, Teladoc Health
Arpana Mathur, MD, MBA, CVS Health
Raquel Mayne, MS, MPH, RN, Hospital for Special Surgery
Anne Myrka, RPh, MAT, Island Peer Review Organization
Edward Pollak, MD, Henry Ford Health System
Nancy Schoenborn, MD, American Geriatrics Society
David Seidenwurm, MD, FACR, Sutter Health
Geeta Sood, MD, ScM, The Society for Healthcare Epidemiology of America
Iona Thraen, PhD, ACSW, University of Utah School of Medicine
Yanling Yu, PhD, Washington Advocate for Patient Safety

NQF Staff:

Poonam Bal, MHSA, Senior Director
Erin Buchanan, MPH, Manager
Tricia Elliott, MBA, CPHQ, FNAHQ, Senior Managing Director
Tamara Funk, MPH, Director
Hannah Ingber, MPH, Senior Analyst
Yemsrach Kidane, PMP, Project Manager
Jesse Pines, MD, MS, MBA, Consultant
Sean Sullivan, MA, Associate

Also Present:

Dan Budnitz, MD, MPH, CAPT, USPHS, U.S. Centers for Disease Control and Prevention (CDC)
Karen Campos, American College of Radiation
Cheng Lin, Acumen, LLC
Mahadevappa Mahesh, PhD, MS, American Association of Physicists in Medicine
Nathan Mazonson, Alara Imaging
Sri Nagavarapu, Acumen, LLC
Simon Rascovsky, MD, Alara Imaging
Francesco Ria, DMP, Duke University
Patrick Romano, MD, MPH, FACP, FAAP, University of California, Davis
Rebecca Smith-Bindman, MD, University of California, San Francisco
Yifei Wang, University of California, San Francisco

Contents

Adjourn	180

Proceedings

(10:04 a.m.)

Welcome and Review of Meeting Objectives

Ms. Funk: My name is Tami Funk and I'm the director supporting the Patient Safety Project. I'd like to welcome you all to our first Patient Safety Measure Evaluation web meeting of 2022.

I'm excited to be here with you today and looking forward to the robust discussions we will have.

I want to thank you for your time and participation, first of all, as I understand that it takes a significant amount of time and effort to review the measures and prepare for today's discussions.

I'd like to extend another thank you to all our measure developers for being on the call today. We also recognize the significant time and effort that goes into the creation, testing, and submission of a measure. And, we want to highlight those efforts and thank you for this important work as well.

Lastly, I appreciate your continued patience and understanding as we continue to meet virtually amid the pandemic. We understand the challenges that accompany virtual meetings. However, our team appreciates your understanding and thanks you for your continued support and participation.

Starting this cycle, the Patient Safety Committee also has two new co-chairs, Dr. Donald Yealy and Dr. John James. Both have served as members of this committee for several years now and are stepping into this new role. I'd like to give both of them a chance to provide some welcoming remarks.

Co-Chair James: Thank you, Tami.

Can everyone hear me?

Ms. Funk: Yes.

Co-Chair James: I just want to add that as I look at the names on this group and the degrees and know of the experience, I'm humbled to be in a role of trying to help this group exercise its intellectual and experience to improve actually patient safety.

I'm a patient safety activist and I feel, through the NQF process, a lot of improvements in patient safety have -- can be done.

One thing that I think about is how we on the NQF Committees can implement change and that seems to come through us making recommendations to the proponents of the measures to improve.

We may endorse their measure, but we may give them a short list of things they might do to improve so that when we see them again in three years, they will have responded to our suggestions.

I pass the welcoming off to Don now.

Co-Chair Yealy: Thanks very much, John, and I welcome everybody for joining today. And, as I wrote in the chat, I thank you for investing your important and valuable time and expertise with NQF, not only getting ready for the meeting, but sharing time today.

John and I get to take over for the first time from Ed and Iona who really led this committee for the past few years, did it wonderfully. And, we have very big shoes to fill, at least figuratively big shoes to fill. I'm not actually certain what size shoe John is and Ed were.

My goal for today, in addition to what John talked about, is that we have developers that have invested a lot of expertise and time in creating measures for our evaluation. Those measures have already undergone some pre-evaluation. We've all seen that, we've all done some of our own review.

I want us to make sure today that we accomplish two

things at the same moment in time, and that's that we give fair and full consideration to each and every measure and the parts of the measure and that we spend the time that's necessary on each measure.

But having said that, I don't want us to spend one minute extra beyond that because I respect your time. And so, I'm counting on us being able to be thoughtful, rigorous, and efficient at the same moment in time. Because everything's important that we're doing including the time that you're investing.

So, part of what John and I will be doing is making sure that the evaluation, the conversation, the exchanges, are not only professional, but they're focused and moving along so that we can accomplish as much as we can, particularly today. I don't want us to have to spend extra time, but we have it carved out if necessary on an additional day.

I think we can accomplish our goals and do them very, very well. It just requires a little bit of structure.

So, if you can work with us and understand that we're starting two good goals at the same moment in time, I'd appreciate it.

We're going to do something a little different during the measures in that any question or dialogue back and forth, we'd like you to put the framework of it in the chat first rather than just open ended questions. That allows you to not only distill down the essence of whatever your insider concern is, it allows everybody to understand it and to be prepared such that the dialogue that follows.

This is not meant to cut off dialogue, it's actually to make it more focused. So, we'll remind everybody about that when we get to the specific measures.

Let me turn things back over to our NQF team.

Ms. Funk: Great, thank you so much for those welcoming remarks.

I want to take a brief moment to review a few housekeeping reminders.

So, as most of you know, we're using the WebEx platform to host this meeting today. I know there are inherent challenges to a virtual platform. So, if you're having any technical difficulties, please let us know. The team is ready to assist you via the chat or by emailing us directly at patientsafety@qualityforum.org.

In the spirit of engagement and collaboration, I encourage everyone to turn on your video if you're able so that we can see each other's faces and bridge some of those virtual gaps.

If you're not actively speaking, we ask that you also please place yourself on mute to minimize background noise and interruptions. To mute, just click on the microphone at the bottom of your screen. And, to unmute, click on the mic again.

As Don said, we also highly encourage everyone to use the chat box feature today to help keep discussion moving and focused and also to use the raise hand feature throughout the meeting.

NQF staff and the co-chairs will monitor discussions and highlight chat comments throughout the call and make sure to call on people with their hands raised.

There is also an option to chat with someone directly if you need that.

For the raise hand feature, the raise hand icon appears in your participant's panel in your video. So, if you click on the participant list, you can hover over your name and the hand icon will appear. Clicking it raises your hand, clicking it again lowers your hand.

Shortly after this, our senior Managing Director of Measurement Signs and Application, Dr. Tricia Elliot will conduct roll call and review disclosures of interests.

It's important to note that we are a voting body and, therefore, need to establish a quorum to vote during our meeting today. If you need to step away from the call, we ask that you please notify NQF team using the chat so that we can maintain an awareness of attendance and quorum members throughout the meeting.

A number of you have notified us in advance of this meeting about short periods of time when you'll be gone. So, if you don't mind putting in the chat when you step away and when you return, that will help us keep a better count on attendance.

Next slide, please?

It's not my pleasure to introduce our project team. I'm Tamara Funk, the director of this project.

Our Senior Manager is Erin Buchanan. Our Manager is Hannah Ingber. Our Associate is Sean Sullivan. And, the extra supporting staff listed here and present on the call today are Poonam Ball, Senior Director, Yemi Kidane, our Project Manager, and Jesse Pines, our clinical consultant.

I am feeling a bit under the weather. I've had a rough week and so, Erin and Poonam have graciously stepped in and will be leading the majority of this meeting. I'll be present as much as possible in the background to support them.

So, now, I will hand it over to Erin.

Ms. Buchanan: Thanks, Tami. Hope you feel better.

Introductions and Disclosures of Interest

So, I'll briefly go over today's agenda. We'll be taking attendance and asking you all to state any disclosures of interest.

Following that, Hannah will provide an overview of the evaluation and voting processes. Hannah will also conduct the voting test.

Poll Everywhere is the online platform we'll be using for the voting process. Last night, you should have received an email with the Poll Everywhere link this morning. I believe the link has also been added to the meeting invite. But, if you can't find it, please send the team a chat and we'll be happy to send it to you again.

After the voting test, I'll briefly introduce our measures under review then hand the discussions over to our co-chairs to facilitate the consideration of candidate measures.

So, following that, you as the standing committee will discuss the measures and discuss each criterion in order and vote on each criterion. The last vote will be an overall recommendation for endorsement for the measure.

Following the discussion of all measures, we'll review related and competing measures for all of those measures that are recommended for endorsement today.

And then, there will be an opportunity for NQF members and public to voice their comments.

And, we'll conclude with next steps and what to expect moving forward and then adjourn.

So, now, I'll hand it over to Tricia to conduct our introductions and disclosures of interest.

Ms. Elliott: Excellent, thank you, Erin, so much. And, thank you all for joining us today and giving of your time to be with NQF for this patient safety fall '21 cycle review.

Today, we will be combining introductions with disclosures of interest. You received two disclosure of interest forms from us, one is our annual disclosure of interest and the other is disclosures specific to the measures we are reviewing in this cycle.

In those forms, we asked you a number of questions

about your professional activities. Today, you'll -- we'll ask you to verbally disclose any information you provided on either of those forms that you believe is relevant to this committee.

We are especially interested in grants, research, or consulting related to this committee's work.

Just a few reminders before we begin, you sit on this group as an individual. You do not represent the interests of your employer or anyone who may have nominated you for this committee.

We are interested in your disclosures of both paid and unpaid activities that are relevant to the work in front of you.

Finally, just because you disclose does not mean that you have a conflict of interest. We do verbal disclosures in the spirit of openness and transparency.

We will now go around the virtual table. I'll start with our committee co-chairs in just a second. I will call your name, please state your name, what organization you are with, and if you have anything to disclose. If you do not have disclosures, please just state that I have nothing to disclose to keep us moving along.

If you experience trouble unmuting yourself, please raise your hand so that our staff may assist you.

So, with that, I will begin with our co-chairs, John James?

Co-Chair James: I'm John James, I run Patient Safety America. I have no disclosures to report.

Ms. Elliott: Thank you.

Don Yealy?

Co-Chair Yealy: I'm Don Yealy, I am a distinguished professor and chair or Emergency Medicine at the

University of Pittsburgh and the Chief Medical Officer for UPMC. And, I have no conflict of interest disclosures.

Ms. Elliott: Thank you very much.

Emily Aaronson?

Member Aaronson: Hi there, I'm Emily Aaronson. I'm the Associate Chief Quality Officer at Massachusetts General Hospital and a practicing emergency physician. And, I have nothing to disclose.

Ms. Elliott: Thank you.

Joel Bundy?

Member Bundy: Good morning, Joel Bundy. I'm the Chief Quality Office for Sentara Healthcare. And, I have nothing to disclose.

Ms. Elliott: Thank you very much.

Elissa Charbonneau?

Member Charbonneau: Hi, I'm sorry, I'm on the phone because I lost my internet connection. My name is Elissa Charbonneau. I am the Chief Medical Officer for Encompass Health. And, I have nothing to disclose.

Ms. Elliott: Thank you, Elissa, we can hear you fine.

Curtis Collins?

Member Collins: Hi, good morning. Curtis Collins, Infectious Disease Clinical Pharmacist at St. Joseph Mercy Health System. I have nothing to disclose.

Ms. Elliott: Thank you.

Theresa Edelstein?

Member Edelstein: Good morning. I'm Theresa Edelstein. I'm one of the Senior Vice Presidents at the New Jersey Hospital Association. I oversee post-

acute care, managed care, and insurance. And, I have nothing to disclose.

Ms. Elliott: Thank you.

Terry Fairbanks?

Member Fairbanks: Good morning, everyone. Terry Fairbanks, I'm Chief Quality and Safety Officer at MedStar Health, professor of Emergency Medicine at Georgetown University. I practice emergency medicine at MedStar Washington Hospital Center in D.C. I have no disclosures.

Ms. Elliott: Thank you.

Jason Falvey?

Member Falvey: Hi, good morning, everybody. Jason Falvey, assistant professor, Department of Physical Therapy and Rehabilitation Science and epidemiology in public health at the University of Maryland School of Medicine. I do receive grant funding from the National Institute on Aging and the Donaghue Foundation related to quality of nursing homes, but I have nothing else to disclose outside of that.

Ms. Elliott: Thank you.

Robert Green? Okay, we'll circle back to Robert. Sara Hawkins?

Member Hawkins: Yes, hi, good morning. I'm Sara Hawkins. I'm the Director of Patient Safety and Risk at Eastern Idaho Regional Medical Center which is an affiliate of HCA. And, no disclosures today. Thank you.

Ms. Elliott: Thank you.

Bret Jackson?

Member Jackson: Good morning, Bret Jackson, President of the Economic Alliance for Michigan and I have nothing to disclose.

Ms. Elliott: Thank you.

Laura Kinney?

Member Kinney: Hi, I'm Laura Kinney. I am Director of Quality at Teladoc Health and I have no disclosures.

Ms. Elliott: Thank you.

Arpana Mather? Okay, we'll circle back. Raquel Mayne?

Member Mayne: Hello, this is Raquel Mayne, Director of Evidence-Based Practice, Research, and Implementation at New York City Health and Hospitals and I have nothing to disclose.

Ms. Elliott: Thank you.

Anne Myrka?

Member Myrka: Hi, I'm Anne Myrka. I'm a Pharmacist and I'm Senior Director of Drug Safety and Chronic Disease Management at IPRO in New York. And, I have nothing to disclose.

Ms. Elliott: Thank you.

Ed Pollak?

Member Pollak: Good morning, Tricia and everyone. Ed Pollak. I'm an Anesthesiologist and I'm the Chief Quality Officer at Henry Ford Hospital Medical Group and nothing to disclose.

Ms. Elliott: Hi, Ed, how are you?

Member Pollak: Hi, Tricia.

Ms. Elliott: Sorry, we worked together before. Nice to see you, Ed.

Jamie Roney? Okay, Nancy Schoenborn?

Member Schoenborn: Good morning, everyone. I'm

Nancy Schoenborn, I'm a geriatrician and Associate Professor of Medicine at Johns Hopkins. I have no disclosures.

Ms. Elliott: Thank you.

David Seidenwurm?

Member Seidenwurm: Very good pronunciation, David Seidenwurm, I'm a neuroradiologist with Sutter Health in California. I'm actually driving to our first off site in person meeting in a really long time. I'm the Network Medical Director and the Quality and Safety Medical Director and, due to conflicts of interest with the American College of Radiology, I'm recused from the radiation safety metrics that are under consideration today.

Ms. Elliott: Excellent, thank you.

Geeta Sood?

Member Sood: Hello, I'm Geeta Sood. I'm an ID physician and the hospital epidemiologist at Johns Hopkins University and also the chair of the Quality Metrics Task Force for SHEA. I have no disclosures.

Ms. Elliott: Thank you.

Iona Thraen?

Member Thraen: Good morning, Iona Thraen, Director of Patient Safety for the University of Utah Health Systems and adjunct assistant professor for the Department of Biomedical Informatics at the U. I have nothing to disclose.

Ms. Elliott: Thank you.

And, Yanling Yu?

Member Yu: Good morning. I'm Yanling Yu, I'm with Washington Advocate for Patient Safety. And, I have nothing to disclose.

Ms. Elliott: Okay. I'd like to call upon Robert Green.

Have you been able to join the meeting?

Okay, I do not see Robert. And, Arpana Mathur? Okay, and, one last call for Jamie Roney or Roney?

Okay. With that, I'll turn it back -- actually, one more thing. I'd like to let you know that if you believe that you might have a conflict of interest at any time during the meeting as topics are discussed, please speak up. You may do so in real time during this web meeting or you can send a message via chat to your chairs or to anyone on the NQF staff.

If you believe that a fellow committee member may have a conflict of interest or is behaving in a biased manner, you may point this out during the meeting, send a message to your co-chairs, or to the NQF staff.

At this point, does anyone have any questions or anything that you'd like to discuss based upon the disclosures made today?

Okay. As a reminder, NQF is a nonpartisan organization. Out of mutual respect for each other, we kindly encourage that we make an effort to refrain from making comments, innuendos, or humor relating to, for example, race, gender, politics, or topics that otherwise may be considered inappropriate during the meeting. While we encourage discussions that are open, constructive, and collaborative, let's all be mindful of how our language and opinions may be perceived by others.

With that, I will turn things over -- turn things back to the team.

Ms. Buchanan: Thanks, Tricia. I'm actually seeing a question here from Geeta that we may want to address first. Can we clarify the role of developer?

Geeta, did you want to elaborate on your question?

Member Sood: Thank you.

I just wanted to clarify what the role of the

developers would be in discussions. Is that something -- my understanding was that they are available to answer clarifying questions, but not necessarily participate in the discussion, is that correct?

Ms. Buchanan: Yes, that's a good way to look at it. Poonam, do you want to provide any clarification? I think that the only difference I think that we used to do is that now we're sort of funneling our questions through our -- the committee and the co-chairs prior to turning to the developer for clarification. But, Poonam, do you want to add anything?

Ms. Bal: Yes, Geeta, thank you for that question.

We are piloting a new structure for involving developers in the discussion. I think Don and John will guide us through that throughout this meeting.

But the core of it is that we're really trying to keep the conversation amongst the standing committee as much as possible and then only going to the developers for clarification or responses that the standing committee themselves cannot answer.

So, the general structure will be the committee will discuss, as Don stated, adding any questions for the developer into the chat and then we will have a facilitated discussion with those questions as well.

So, bring them to the standing committee first to see if there's anyone who can clarify and then go to the developer. I will say, the developer still has an opportunity to raise their hand or put a comment in the chat if they feel that, you know, something -- it would be important to clarify something sooner than later, they still have that opportunity.

Does that make sense? Any other questions?

Co-Chair Yealy: So, Geeta, I think what it will be is that there'll be more structured back and forth between the developer and the committee.

And, that's why having a typed in question or clarification first will allow much more focused, not only preparation on the developer's side, but for us to know what needs to be addressed and not addressed.

Ms. Buchanan: Okay, great. It looks like that clarified things for everyone.

If there are no other questions, I'll turn it over to Hannah to discuss our overview of the evaluation process and voting process.

Overview of Evaluation Process and Voting Process

Ms. Ingber: Thanks, Erin.

So, we'll now transition to a brief overview of the evaluation and voting process.

Your role as a standing committee is to act as a proxy for the NQF multi stakeholder membership. As the Patient Safety Committee, you not only oversee the portfolio of patient safety measures, but you work collaboratively with NQF staff to provide recommendations for endorsement of measures based on our CDP evaluation guidance.

You're also tasked to respond to comments that are submitted during our public commenting period.

Today, you'll be asked to evaluate measures against each criterion and subsequently make recommendations according to your evaluation.

Next slide, please?

So, we'll go over some meeting ground rules. The first one is, there's no rank in the room. And, to clarify, this is a shared space of interdisciplinary multi stakeholder committee members. Every voice is important and we want to emphasize that each committee member holds equal value on this call and in the broader scope of the work.

We ask that you remain actively engaged and actively participate in the call. And, be prepared, having reviewed the measures beforehand.

Please base your evaluation and recommendations on the measures on the measure evaluation criteria and guidance.

Keep your comments concise and focused. Be respectful to allow others to contribute, and please share your experiences, and, of course, learn from others.

Thank you.

So, this slide describes the process by which we'll conduct today's measure discussion and evaluation.

Each measure with discussion will begin with a brief developer introduction. And then, facilitation will be led by the co-chair and discussion will be stewarded by our assigned lead discussant and supporting discussants.

Thank you, again, discussants, for your leadership today.

The lead discussant will briefly explain information on the criterion, emphasize notable areas of concern and note the preliminary staff rating, if needed.

Full committee discussion will then commence, followed by the criterion vote.

The process will be repeated with each subsequent criteria. And, again, developers will be available to respond to questions at the discretion of the co-chairs.

And, as a reminder, we'll vote on each criterion as we discuss it.

Moving to the next slide, please?

These are -- measures are evaluated for their

suitability based on a standardized and main and subcriteria in the order depicted on the screen.

So, first, we have importance to measure and report which examines the extent to which the measure focus is evidence-based and important to make significant gains in healthcare quality where there is variation in or overall less than optimal performance.

We then have scientific acceptability which has subcriteria of reliability and validity which examines the extent to which the measure produces consistent and credible results about the quality of care when implemented.

Feasibility looks at the extent to which the specifications require data that are readily available or could be captured and implemented without undue burden.

Feasibility and use are two subcriteria that examine the extent to which the measure is being used for both accountability and performance improvement to achieve the goal of high quality efficient healthcare.

And, as a reminder, uses must pass for maintenance measures, of which we have one today.

Then, if a measure meets the above criteria and they are recommended for endorsement, we will then commence with a comparison to related and competing measures discussion.

The measures are compared to address harmonization and/or selection of the best in class measure.

Okay, the breakdown of the main endorsement criteria and some criteria is, again, listed here. And, as I mentioned before, votes will be taken after the discussion of each criterion. So, please make special note of the must pass nature of several of these criteria and that it differs between new and maintenance measures.

If the measure progresses to the last criteria, the overall suitability for endorsement will be the last vote.

If a measure does not pass on a must pass criteria, we will not vote on overall -- we will stop voting on all the remaining criteria and will not vote on overall suitability for endorsement.

Okay, NQF staff will provide a brief overview of the related and competing measures and will invite the committee to weigh in with any further commentary.

It is important to reiterate that, again, if measures fail on one of the must pass criteria, we will not proceed to any additional discussion or voting on the subsequent criteria.

However, if consensus is not reached, discussion will continue to the next criterion but a vote on overall suitability will be deferred to the post comment meeting.

Next slide, please?

Okay, this slide is text heavy, but I'll explain everything.

In order to conduct live voting today, the standing committee must achieve and maintain quorum, which is 66 percent of attendance of its active participants. That is, for our committee, 16 of 23 members.

We have 20 members present today, so we have reached quorum. Thank you, everyone.

And, to clarify the chart in the second half of the slide, it displays the margins within which voting outcomes are indicated. So, a measure that does not reach consensus will move forward, again, but measures that do not reach consensus where fewer than 40 percent of the committee votes yes, will not move forward.

Again, a yes vote is the total of high and moderate votes or pass votes based on the number of active and voting eligible standing committee members who participate in the voting activity.

If a measure is not recommended for endorsement, it too will proceed to the draft report commenting period, but the difference here is that the committee will not be called upon to revote on the measure unless the committee decides to reconsider their recommendation based on either comments from the draft report commenting period or a formal reconsideration request from the developer.

Next slide, please?

Okay, as stated on the previous slide, 16 active committee participants must be present in order for the committee to vote. And, we also add that that baseline 50 percent of active committee members must be present in order for the call to be held at all.

So, this is where attendance plays a significant role. If, again, if at any point you need to leave or step away for a minute, please just place a note in the chat to let us know about this so that we can monitor attendance accordingly.

In the event that attendance drops below quorum, we will resume discussion respective to the measure at hand, but we will defer voting activity to an offline voting survey that will be sent to the committee after the call.

If a committee member leaves the meeting and quorum is still present, the committee will continue to vote on the measures and the committee member who left will not have the opportunity to vote on that measure or criteria, depending on when they return that were evaluated by the committee during their absence.

Next slide, please?

Are there any questions about the evaluation process or the voting process that we went over?

Okay, hearing none, oh, I'm just checking the chat, thank you.

Okay, I don't see any in the chat, but please feel free to speak up if you do have a question.

Voting Test

All right, we'll proceed to the voting test using the link that you were provided in your email this morning or last night. As a reminder, this voting test is just for the standing committee members. And, we'll be looking for a total of 20 in our responses.

Ms. Buchanan: And, I just want to jump in really quick and let everyone know that the voting link is only for standing committee members, so please do not -- any other participants on the call today, please don't click that link or participate in the poll.

We will also be double checking the poll in the background to insure that committee members are the only ones participating.

Ms. Ingber: Okay, we're looking for just three more.

Member Bundy: Yes, this is Joel Bundy. Mine did not pop up so I'm going to sign back into the polling.

Ms. Ingber: Okay, thank you.

Ms. Bal: We do find that if you refresh your screen, that often helps with keeping the voting -- open up the voting for you as well.

Ms. Ingber: Okay, I see 20. So, I'll now close the poll and share the results. So, on the question of whether you like broccoli or not, 17 members said yes and 3 members said no.

I will now turn it back to Erin to go through the first measure.

Ms. Buchanan: Great, thank you. Thank you for pulling up the slides, Sean.

All right, so, for our measures under review, next slide, please?

Measures Under Review

So, we have five measures under consideration today, one maintenance measure, and four new measures.

The maintenance measure is Number 0689, Percent of Residents Who Lose Too Much Weight (Long-Stay). The developer is the Centers for Medicare and Medicaid Services.

The first new measure is 3636, Quarterly Reporting of Healthcare Vaccination Coverage Among Healthcare Personnel. The developer is the Centers for Disease Control and Prevention.

Following these, we have three new measures by the University of California San Francisco and Alara Imaging.

3633e is Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level), 3663e is that same measure of the Clinician Group Level and 3663e is the same measure at the Facility Level.

A quick review of the Scientific Methods Panel, this is a group of researchers, experts, and methodologists in healthcare quality and quality measurement. The SMP reviews complex measures and the panel's comments and concerns are provided to the developer so that they can provide further clarification and update their measure submission form to strengthen the measures that are to be evaluated by the standing committee.

So, when you got the measure evaluation form prior to this meeting, you will see the SMP's review contained within it along with staff's review.

Next slide?

So, four patient safety measures for review by the SMP this cycle, three measures passed the SMP for reliability and validity, 3662e, 3663e, and 3633e. So, those three passed both reliability and validity at the SMP.

0689 was also reviewed by the SMP and passed on reliability but was consensus not reached for validity.

And, I'll pause for a second, Geeta has a question regarding the SMP's review. So, Geeta's asking, as I understand the Scientific Committee is evaluating use agnostic, so it is truly incumbent upon us to review the context of each measure, correct?

Yes, so, the Scientific Methods Panel is only looking at the testing pretty much. So, there's a whole category around use and usability which you all will discuss.

Poonam, I don't know if you want to add anything to that.

Ms. Bal: No, I think you -- what you said is right. We will talk about use but we won't be considering intended use when considering the merits of the measure.

Ms. Buchanan: Right.

And then, Iona, for 3636, that is a process measure so it was not reviewed by the SMP. So, we will be looking -- we will be looking at scientific acceptability and the committee's comments around that during the discussion today.

All right, any other questions about the SMP?

Okay, so, let's see, can you go over -- okay, so, we'll now begin our review of the -- of our fall 2021 measures.

Just a quick recap of our process today, our co-chairs

will start us off by introducing the measure. The developer will then have three to five minutes to provide a brief overview of their measure. The co-chairs will then hand things over to our lead discussants who will summarize the measure and review committee and public comments.

During the discussion, any questions from the standing committee for the developers will be noted by the NQF team and co-chairs. After sufficient discussion, the co-chairs will pause the discussion and allow the developer to clear up any additional concerns that weren't addressed by the committee.

The discussion will then turn back to the standing committee. If at any time during the discussion the developer team would like to address any inconsistencies they hear or clear up some confusion, we ask that you please use the raised hand feature and put your request in the chat.

The co-chairs will call on the developers to speak at the appropriate time. And, as a reminder, the full committee will be discussing and voting on each of the measure criteria. So, for every single measure.

And now, I think it's time to hand it over to John to start the discussion on 0689.

Next slide, please?

Consideration of Candidate Measures 0689 Percent of Residents Who Lose Too Much Weight (Long-Stay) (Steward/Developer) (Centers for Medicare & Medicaid Services (CMS))

Co-Chair James: Great, thank you, Erin.

This measure has to do with residents who lose too much weight while they stay in a nursing home. It captures a certain percentage of long-stay residents whose weight falls below five percent over a period of 30 days and there is no need from a physician or an indication of the need for this level of weight loss.

Also, ten percent or more over six months is also a consideration.

I'm now going to turn it over to CMS, our developer, to discuss their measure.

Ms. Lin: Hi, John, this is -- his this is Cheng from Acumen. Can people hear me?

Co-Chair James: Yes.

Ms. Lin: Okay. Thank you for the opportunity to further introduce this measure. As John described, this measure reports the percentage of long-stay residents in a nursing home who had a weight loss of five percent or more in the last month, or ten percent or more in the last six months, which is not a result of physician prescribed weight loss regimen, residents under hospice care or with a life expectancy of less than six months are excluded from measure calculation.

Long-stay residents are those who have stayed in the nursing home for over 100 cumulative days in the current abstract.

The measure is based on data obtained through the MDS data set or Minimum Data Set. The MDS is an assessment tool that is required to be filled out routinely for care planning and resident status checking. So, there is no extra burden for nursing homes to collect measure information.

This measure has been endorsed several times. It was last endorsed in 2015. This measure has a long history in the nursing home quality initiative program which started in 2002.

All intended and excessive weight loss is an important measure for nursing homes. Research shows that weight loss is associated with higher risk of hospitalization and increased mortality.

The measure is publically reported through the Care Compare website and provider data catalogue.

The Care Compare site, or previously known as Nursing Home Compare, is a tool for patients and families to learn the quality of care at different facilities and select providers.

Provider data catalogue is a site commonly used by researchers. In addition to use by the patient and research community, confidential feedback reports are also available to providers through CASPER. And, providers can access patient level information to identify areas for improvement and take action to work on weight loss monitoring and prevention.

This measure has demonstrated improvement over time. The four-quarter average rate decreased from six percent in 2014 to five point three percent in 2019. This percentage change corresponds to over 8,000 beneficiaries given the number of 2019 stays.

This measure has passed SMP reliability voting, but consensus was not reached on validity.

On the data elements underlying the measure demonstrates strong validity through the risk study. The result is still highly relevant because the data element in the MDS assessment hasn't change since it's been conducted real study.

When reviewing the committee worksheet, we noticed that one question about validity that came up is the lack of risk adjustment. As Acumen explained in the appendix of the measure submission, the measure is intentionally not risk adjusted.

Acumen estimated risk adjustment models using the risk factors suggested by the Scientific Method Panel and a series of other potential risk factors.

The results demonstrates that the most predictive covariant are items that are under the control of the facility and, hence, inappropriate for inclusion in risk adjustment.

The inclusion of other covariant has minimal impact

on the rank of provider scores. The deceased statistic of only .51 which is not close to a coin toss.

This model also leaves almost all facility decile change rankings unchanged. More specifically, 97 of the decile ranking remains in the same decile.

Moreover, there is a strong clinical argument to not risk adjust for these covariant given the incentive this would create.

I'm happy to elaborate more on these covariant in further discussions.

These findings support the decision to not risk adjust this measure.

This ends my introduction and I'll turn it back to the NQF staff group.

Co-Chair James: Well, thank you.

I think there was a clarification in the chat that CMS is not the actual developer, that Acumen, LLC is the developer, for whatever that name fosters.

Meanwhile, I think it's time to go on now to the discussants for this measure. The lead discussant is Dr. Seidenwurm. David, would you please go ahead?

Member Kinney: I believe I'm taking his place today, Laura Kinney.

So, regarding importance to measure report, the evidence is a must pass criteria. So, during the last review cycle, the standing committee agreed that the evidence was strong for this important outcome measure.

Concern with the lack of data on disparities and the lack of observable improvements were raised.

The developer stated that the lack of change in this measure may indicate that nursing homes are not improving in this area, highlighting the need for

continued public reporting on it.

The committee noted a greater effort to delay patient admission to long-term care. This, in turn, caused the populations in nursing homes to be increasingly frail which leads to difficulty in maintaining nutritional status.

The developer did present new evidence. They offered updated research studies to further demonstrate the negative association between excessive and unintentional weight loss and other health outcomes, including risk of hospitalization and increased mortality.

The developer also presented the impact of COVID-19 on both residents who contracted the virus and those who did not. Small percentages of residents experienced unintended weight loss in both infected, which was ten percent, and non-infected populations, seven point five percent in the study from 2021.

They also presented updated evidence of several actions that nursing home staff and facilities can take to prevent unintended weight loss.

Co-Chair James: Are there further comments?

Member Kinney: No, I'm finished.

Co-Chair James: Okay, thank you.

The other discussants are Raquel Mayne and Theresa Edelstein. Do you have comments to add?

Member Mayne: This is Raquel Mayne, I have nothing to add.

Member Edelstein: Same, I don't have anything to add to Laura's comments.

Co-Chair James: Okay.

Can someone provide the feedback from the committee members prior to the meeting on this

measure? Were there concerns raised in that feedback?

Member Kinney: No, there were not.

Co-Chair James: No concerns at all? Wow, that's a clean state.

Member Kinney: Not in the evidence component, no.

Co-Chair James: All right.

In the chat, are there questions we need to deal with?

Co-Chair Yealy: John, I'll take this one question from Geeta about the modifiability, essentially the causality of the measure and how that might play into the evidence behind it.

This is, again, it's very hard on these on these questions to stay inside of a lane, whether this is really purely an evidence question or, you know, drifts into another part of the evaluation is always one of the challenges. And, I don't -- after eight years, I've not figured this out, how to ask the singular pure question.

And so, this might be something I think the developers tried to share their thinking behind, why they didn't adjust for certain features, in other words, it does have an evidentiary component to it. But it's a fair question.

Co-Chair James: Yes, are there any further comments on the lack of risk adjustment?

Co-Chair Yealy: So, I would say, thinking now as a committee member, I would -- while I understand that different features may link to a varying body mass kinetics or trajectory, I'd have to ask myself, does it really matter why it is? Because, at the end of the day, even if you have a condition that makes it more or less likely, what stands undisputed is that unintended weight loss is not a good thing. It's a sign of a change in trajectory and a negative one.

And so, whether we're talking about the individual or the facility level, I think you end up at a very similar spot.

Member Kinney: The Scientific Methods Panel spoke --

Ms. Bal: I'm sorry, Laura, to jump in here. So this is Poonam from NQF.

If we could hold off on discussing risk adjustment or the specifications of the measure until we get to those areas, that would be great. Right now, we're just focusing on, is there evidence to support this measure.

But thank you, Laura, for jumping in. But just, we should discuss evidence first and then move forward.

Member Kinney: Right.

Ms. Bal: Thank you.

Co-Chair James: Thank you for keeping us on track.

I think we have no direct questions then for CMS, so I think we should go to the voting. Are we ready for that?

And, I think Hannah's going to measure -- conduct that, if I understand my agenda here.

Ms. Buchanan: Yes, you are. And, I do want to point out that there is -- we sent out a new link. So, if you all could check your email, there should be a new link that you'll use for this measure.

Co-Chair James: New since when? This morning?

Ms. Buchanan: Yes, since this morning.

Ms. Bal: Yes, so just for clarity, we wanted to make sure that the link stayed amongst the standing committee, so we did create a new one and that was sent probably 30 minutes ago, 20 minutes ago, something like that.

Ms. Buchanan: More like 10, yes.

Co-Chair James: So, our test of the platform was null and void?

Ms. Bal: Yes, unfortunately. But we'll see if there's any -- the platform's the same, the link has just changed. So, hopefully, that doesn't cause any issues but if you are experiencing any issues, please message us and we'll make sure to resolve that.

And then, as a reminder, if you are not able to vote, it's always best to refresh your screen. If you go to the old link, it will not work. So, no -- so, you'll know right away if you're in the wrong --

Ms. Buchanan: And, please do not share the new link in the chat.

Co-Chair James: So, we should do a test, I guess, with the new link to see who can get on it.

Co-Chair Yealy: I think our test will be just to vote because they can tell who voted. We'll just vote on this part of the measure now.

Ms. Buchanan: Yes.

Ms. Ingber: That's correct. And, just for formality sake, I'll say that voting is now open on importance to measure report for Measure 0689. Your options are pass and do not pass, as this is an outcome measure.

I'm seeing 17 results and I'll pause just since the voting link changed, but we are expecting 19.

Member Sood: Sorry, I know this is supposed to be on the chat, but I can't get into the link. The email that was just sent out today.

Ms. Bal: Geeta, I'll privately chat you the link so you have it. Does anyone else need a private chat with the link? And, I think we've got our 20.

Ms. Ingber: Yes, voting is now closed on Measure 0689 for importance to measure and report one-A, evidence. I'll now share the results.

So, for Measure 0689, we have 18 counts for pass and 2 votes for do not pass. Therefore, the measure passes on evidence. Thank you, everyone.

Member Kinney: Okay, our next voting opportunity is for the opportunity for improvement. This is a must pass criterion.

The developer did report an analysis of four-quarter facility level data of 14,274 facilities between quarter one and quarter four of 2019.

The mean performance was 5.2 percent with a standard deviation of 3.1 percent and a range of 1.6 percent to 9.2 percent.

The developer also notes that the IQR of 3.9 percent and the small number of facilities with perfect scores, 2.6 percent, indicates there's still room for improvement.

Disparities were noted, population older than 85 years is at a slightly higher risk of losing too much weight, 5.9 percent when compared to younger residents, 5.1 percent.

The white population is a slightly higher risk of losing too much weight, 5.4 percent, when compared to non-white populations at 5.2 percent.

Pre-evaluation comments included, quote, the evidence suggests that unintended excessive weight loss is associated with increased mortality in the elderly population and can be an indicator of quality and safety care at long-term care facility. The evidence highly supports this outcome measure which can be used to direct quality improvement for long-term care as well as, quote, the evidence presented for this measure is directly related to the measure and they have partnered with it

improvement actions that could be improvement actions that could be levers to affect poor performance.

They have also presented new evidence supporting the measure's focus.

And, I am -- any questions?

Co-Chair Yealy: I think you're muted, John.

Co-Chair James: Don, how do we want to work this from the chat? There's a question in there.

Co-Chair Yealy: I don't think -- I think the chat is part of an exchange on the previous issues. I don't think it has anything to do with the gap now. So, what I see is that there are no focused on this part of the evaluation.

Co-Chair James: Okay. So, at this point, there aren't any questions, then.

Right, well, let's go ahead and vote then.

Ms. Ingber: Okay, I will open the vote, just give me a moment.

Okay, voting is now open for Measure 0689 on important to measuring report performance gap. Your options are high, moderate, low, or insufficient.

Just waiting on one more. I'll just wait a few more seconds and I'll -- then I'll close the poll. We're waiting on one more person, but we do have 19 votes.

Okay, I will close the poll and share the results.

So, for 0689 1-B, in performance gap, we have 4 votes for high, 13 votes for moderate, 2 votes for low, and 0 votes for insufficient. Therefore, the measure passes on performance gap.

Member Kinney: Okay, and we're going to begin to discuss reliability and we'll cover both specifications

and testing together.

The numerator is the number of long-term nursing home residents with a selected target assessment indicating a weight loss of 5 percent of the baseline weight in 30 -- last 30 days or 10 percent in the last 6 months.

The denominator includes all long-stay residents in the nursing home who have a target assessment OBRA, PPS, or discharge during the selected corridor and who do not meet exclusion criteria.

The denominator exclusions include target assessment is an OBRA admission assessment or a PPS 5-day assessment; two, having a prognosis of a life expectancy less than 6 months; three, receiving hospice care; or, four, the weight loss item is missing.

Only 1,500 episodes in the 2019 long-stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months which accounts for 0.04 percent of the total episodes.

On reliability testing, the Scientific Methods Panel is satisfied with the reliability testing for this measure.

The developer conducted performance score accountable entity level testing for the 2021 submission using 2019 data from the MDS 3.0. The developer's national test of MDS 3.0 items used data from all long-stay residents who met denominator inclusion criteria for this measure in facilities with a sufficient sample size and reported this measure between 2019 Q-1 and 2019 Q-4.

Nine hundred, thirty-two thousand residents met the denominator inclusion criteria for testing in these facilities. The developer used inter-rater reliability testing to examine the agreement between assessors.

The developer conducted split-half reliability testing and signal to noise testing reporting that the split-half correlation of this measure was positive and the relationship was moderate suggesting there is evidence of internal reliability.

The developer reported an average signal to noise reliability score of 0.76 using the facility scores based on 2019 data.

Again, the Scientific Methods Panel was satisfied with the reliability testing for this measure.

Are there any questions?

Co-Chair James: Okay, I see no questions. There was a comment in the chat that was supportive of the Scientific Panel conclusion.

I think we ought to go ahead and vote and the voting is going to be whether to accept the Scientific Panel's thumbs up or reliability.

Ms. Bal: John, could we just pause for a second before we jump into voting? I just want to make sure that, even though there was no comments in the chat, that the other discussants didn't have any concerns or if anyone from the standing committee wanted to verbalize any questions or concerns they might have before we get to the voting.

Member Mayne: I have no concerns with reliability.

Ms. Bal: Okay, great.

You were on the right track, John. I just wanted to make sure that we at least paused to make sure if someone wanted to speak up they had the opportunity.

Thank you.

Co-Chair James: Pauses are of different lengths.

Ms. Bal: Yes.

Co-Chair James: I'm a bit like Don, let's keep the ball rolling.

So, are we ready to vote now, folks? Speak now about reliability or hold your peace. Okay, let's vote.

Ms. Ingber: Okay, voting is now open for Measure 0689 on whether you would like to accept the Scientific Methods Panel's rating for reliability. Your options are yes and no. And, we're looking for 19 votes.

Voting is now closed on Measure 0689, whether to accept the Scientific Methods Panel's rating for reliability.

We have 19 votes for yes and 0 votes for no. Therefore, the measure passes on reliability.

Co-Chair James: Okay, Laura, if you would go on to deal with validity now, please?

Member Kinney: Oh, I'm going to pass that to Theresa.

Member Edelstein: Laura, thank you for carrying the ball so far and don't completely bail on me here.

Member Kinney: Oh, no, I'm not.

Member Edelstein: The validity and risk adjustment conversation, I think, will -- could be an interesting one.

As was already mentioned, the Scientific Methods Panel did not reach consensus on validity. At the patient encounter level, testing on the patient or encounter level testing element, there was a high degree of correlation or agreement between the standard nurse assessment to facility nurse assessment of weight loss at 0.918. And the kappa for gold standard nurse assessment to facility nurse assessment of the six month prognosis item was 0.964.

However, at the encounter entity level, multiple members of the panel of the SMP had concerns about the decision not to risk adjust the measure. And, they have suggested that we should discuss whether certain MDS items, and there's a long list of them with different diagnoses such as Alzheimer's, cancer, depression, cardiac disease, and so on as well as eating dependencies might warrant a risk adjustment.

You heard Acumen describe how they addressed that concern. And, they used a convergent validity approach and did see that there was low to moderate correlation between some of those diagnostic items on the MDS and weight loss.

But none of that changed where the facility score ended up on this particular measure.

So, there were among this group, several reviewers did have comments about whether the developer had adequately explored the need for risk adjustment and there were some comments about literature review showing that there are potentially addressable risk factors for unintentional weight loss in long-term care facility residents, citing some of the same conditions that the SMP raised.

So, I think I will stop there, see if there are other comments or questions to kind of build on this conversation that's in the reviewer's comments as well as the SMP.

And, Laura, if you have anything you would like to add?

Member Kinney: No, I think this will be an interesting discussion.

Co-Chair Yealy: John, it looks like our chat question is right back to this again which it's a concern existing in a few areas but probably more squarely here.

It has to do either confounding or multiplication of

observations in different subpopulations. And, obviously, the Scientific Panel had questions about that. And, obviously, the developer has actually already considered that and offered some of a response.

And, now, Geeta has a comment and others have about how this might be handled moving forward.

Co-Chair James: And, what -- how would we handle that as a committee?

Co-Chair Yealy: The question here is dialogue back and forth, do we need more conversation among ourselves in order to make a decision or do we need any further input from the developer before we vote?

And so, that's what I would turn back to the group. The conversation is back -- it's very clear to me from reading the comments that the concern is confounding here. In other words, different populations within any of these care facilities may have different trajectories and frequencies of the phenomenon being measured, and that's unintended weight loss and how could that impact the assessment at any particular level?

And, the developers have done some analysis towards that about what the impact would be. I don't want to speak for them, but as I read it, it essentially says, unintended weight loss, no matter which global or subpopulation you're in, is a marker for health, a negative marker for health and outcome.

And, while there may be different trajectories, adjusting for it isn't necessarily going to make the measure more or less effective. It may make it more complicated.

I think what I see being asked back is, could we at least report amongst those different groups for the next go round?

And, I see Yanling has her hand up. Maybe I'm

mischaracterizing it, of the concerns and the potentials.

Co-Chair James: Yes, go ahead, Yanling.

Member Yu: Thank you, thank you, Don.

My thought is as you alluded that the developer actually looked at the potential for risk adjustment and I agree with the conclusion. And, they didn't say significant impact of the score from the facilities.

On another side, I think, you know, we all know that patients in a long-term facility have a different medical condition or disease so that would impact either how they absorb food and process food. And, you would have impact there, different rates or a different process to lose weight.

But in terms of quality and how the facilities to pay attention to help those elderlies to not losing weight, even though with those health conditions, I think this measure would help them. And, in particularly, probably pay more attention on those vulnerable elderlies who have special, you know, need for a more, you know, assistive to help them maintain a healthy weight, even with their, you know, more serious health problem.

So, I think the risk adjustment, I agree with the developer on that and not using the risk adjustment on the patient population.

Thank you.

Co-Chair Yealy: Jason, you have a hand up, too?

Member Falvey: Yes, I don't know if I was next in line, I think Iona had her hand up first, but I'm happy to quickly share my comments to keep this moving along.

I certainly appreciate the conversation. I think this is a really important measure, so I don't want that to get lost in my comments as a geriatric physical

therapist working with this population, you know, researching and clinically. It's certainly something I could have a lot of impact as clinicians.

There are some situations, I think, when we test methodologically risk adjustment, what we saw with the effect on the average facility and, you know, thankfully, that some of these things, you know, were not affecting the measures and not, you know, moving the average facility from one rank into another.

But my concern is facilities that are highly specialized like long-term care facility that take mechanically ventilated patients, for example. They are going to be very concentrated in a very small number of facilities.

Similarly, patients with other complex feeding needs that potentially might not be captured in MDS. They tend to cluster or concentrate in very specialized facilities. And, I think that there's probably ways that the developer can see facilities that care for high concentrations of high risk patients, whether or not they're being disadvantaged on this measure as a way to capture risk adjustment.

Because, like I said, what the approach the developers took told us the average effect on the average facility, which is important, but maybe not sufficient for some facilities who care for a high concentration of very complex patients.

And, it may be a moot point, but it's certainly something as long-term care facilities become more specialized and patients become more frail and complex, I think we do have to be mindful of not disadvantaging facilities that might serve a very special population.

Co-Chair Yealy: Thank you, very well said.

Iona, I see you have a hand up?

Member Thraen: First of all, get clarification, this is a maintenance review for this particular measure, which means it's been in place for, this is what, it's sixth year or third year?

Member Edelstein: Sixth.

Member Thraen: Sixth year? Has this issue of risk adjustment come up in the past and fed back to the developer in past reviews? Do we know that?

Member Kinney: Basically, they have researched and spoke to that in their research. So, they are aware of it.

Member Thraen: They're aware of it, but they've not incorporated it at this point in time?

Member Kinney: They have chosen not to do that.

(Simultaneous speaking.)

Member Edelstein: They did run models on some of them.

Member Thraen: Correct.

Member Edelstein: Sorry, were recommended by the SMP. And, I think some of the other comments about other variables that may need to be looked at are more operational in nature, things like staff tenure, early detection of weight loss, eating environment. Those things aren't measured on the MDS.

Some of the staffing related metrics that are being suggested are beginning to be reported on Care Compare because CMS has begun to collect that information through the Payroll Based Journal data portal. And, they may prove useful for future analysis and consideration for other risk adjustment approaches.

But, you know, speaking as a former nursing home administrator, I, you know, I think that the -- we should not lose the forest for the trees and really

understand how important it is to pay attention to nutritional status regardless of underlying clinical condition or disease because it's a quality of life issue.

Food is one of the most important considerations for residents and their families. And, enjoyment of food, in particular, in the social nature of it contributes greatly to the mental status and overall life of the resident in these facilities.

So, I just don't want us to lose that piece of the perspective.

Member Thraen: So, I -- just for clarification, so it doesn't sound like, even though they've investigated that this review committee in the past, in the last review process, the first three years, made strong recommendations to include risk adjustment historically.

So, we're now at a place six years later where refinement of the measure and additional evolution of the measure is really what the question is. That's kind of how I'm seeing it.

All right, thank you.

Co-Chair Yealy: Elissa, you have a hand up?

Co-Chair James: I think you're on mute.

Member Charbonneau: Sorry, I wanted just to echo the comments that Theresa made.

As someone who works in post-acute care, I think that it is really, really vital to pay attention to nutritional status of these patients. And, if this measure is accomplishing that, even though it may not be perfect, it is better than not having the facilities accountable for this because nutritional status has so many other implications for these frail, vulnerable populations, including wound healing and other things.

And, my concern would be that if we drop it, that we will lose that attention towards that very important aspect of their care.

Co-Chair James: Don, can I -- I think I'm the only question left.

Co-Chair Yealy: No, I think there's -- we have one left. Nancy?

Co-Chair James: Oh, okay. Nancy, go ahead.

Member Schoenborn: Okay. I just wanted to echo, I think some of what's been said. You know, as a geriatrician, although there are conditions that may be more at risk for developing unintentional weight loss, I think there are studies demonstrating that there are interventions to prevent those.

I don't want us to have a mindset of accepting somehow more weight loss, you know, in dementia patients or in older patients, for example, older age itself as a risk factor. Right? That those are actually specifically the patients that need this measure to help, I think, highlighting their added needs so the facilities can address them.

Co-Chair James: Can I go ahead now, Don?

Co-Chair Yealy: Yes, and I'm not sure that there's a specific set of questions to address to the developer. I mean, I think this has all been discussed before and they have now our thoughts and understandings about from an analysis perspective what might be helpful. I'm not sure if there's anything else to ask them to clarify right now or to respond to.

Co-Chair James: I have a question about, what I read in medical journals is that there's a growing tendency for nursing homes to become for profit. And, I think we all know that may or may not be a good thing, but it could impact the weight loss thing.

I'd like to see that as part of the risk adjustment, if it can be done feasibly. I don't know how hard that

would be to do.

It seems to me that our choice here, and you NQF guys can give us guidance, I think there's a strong enough opinion that the developer needs to go back and do a thorough risk analysis and adjustment, if necessary. But we need them to come back in three years with really convincing data that they've dealt with the risk adjustment question.

Now, the question is, what do we do with the validity? It's a must pass criteria. So, do we let this pass and give advice to Acumen or how do we proceed?

Co-Chair Yealy: I think we have to make the decision right now, yes, on what's in front of us. And, then, assuming we pass, they can heed the comments or at least accept them. If it doesn't pass, that's a whole separate other issue, right?

What I don't think we can say is it needs to be changed and bring it back to us.

Co-Chair James: Not bring it back, Don. But in the next cycle, three years, we would expect them to come back and show us that they've dealt with the concerns that we have now.

Yanling?

Member Yu: Yes, thank you, John.

I just want to quickly share a story that I personally know. There was an elderly patient in the long-term care on the ventilator. And, apparently, the patient definitely was malnutritioned. But families stay in a long-term care, they will not want their loved ones to say, you were right about this.

They say, we have this, you know, terminal disease, and then, therefore, they don't have to follow the best care practice to help them, provide the necessary nutrition to help maybe get better. Nobody knows. You know, someone would turn wrong.

So, I just want to share this little tidbit of what is important to a patient and families when it comes to taking care of them when they have a terminal illness and no someone say it's a goner. You don't have to, you know, be right at all.

Co-Chair James: Okay, thank you, Yanling.

I think Emily has her hand up.

Member Aaronson: I do, thanks so much. And, thank you, Yanling, for sharing that comment. I think it's so important that we continue to ground these in those individual patient stories which are really powerful.

I think the two things that I would offer is, one, I completely appreciate the perspective around not wanting to accept differences in care for certain populations and this idea that if there are some populations that are at higher risk for having difficulty maintaining weight then we shouldn't just accept that.

And, I think it's important that there's some evidence that these are modifiable.

And so, I would ask very specifically that the developers include in their next submission evidence around those specific subpopulations which have been raised here today and any evidence that exists that these are actually modifiable.

Because I think in the absence of that, then you do really get to that concern about negatively penalizing facilities that are disproportionately caring for those populations without actually any proof that they have levers to change that.

And then, I think that the second piece is I think more for the NQF staff, and for us, to understand if there's a more systematic way for us to track the outcomes of these conversations in future. Because I think, you know, on resubmission, it is hard for us to recall, many of us weren't even there, what those sort of

asks of the developers are.

And, I worry that there could be measures that just continue to eke by on conversations like this without any of these things being systematically addressed.

Co-Chair James: I think that's a question for the NQF staff.

Ms. Bal: So, I Just wanted to jump in following the hand raising opportunity. So, we do track those recommendations. We will put them in the draft report and then that is brought into our analysis of the measures in the next round.

So, you know, as long as it's relevant we'll bring it up in the PAs that we share with you.

So, I think that's what I just want to highlight. So, if, you know, again, as John mentioned, we should vote on the measure as is, on the quality it's providing. If there's areas of improvement that we want to highlight, we can definitely do that and we'll make sure it's tracked in the draft report and we'll ultimately pull into the next review cycle.

And then, sorry, while I have the opportunity to speak up, I have been noticing there's a robust conversation in the chat which is great. But if we could try to keep the conversation, you know, verbally that would be better so it's all tracked and in our transcript and recording.

While I think a lot of these discussions are happening and the answers are being provided, we just want to make sure that anyone who, you know, might listen to this later, they can get the full picture of what the discussion was.

So, please feel free to put questions in the chat, but let's try to bring that back into the actual conversation and so we can make sure that everything is publically noted.

Co-Chair James: Theresa, you have your hand up?

Member Edelstein: Yes, I just wanted to make note of a comment in the appendix that was submitted by Acumen, it's really at the very end where they acknowledge that the necessity and appropriateness of risk adjusting the measure should be further evaluated with a focus on the cancer item that was added to the MDS in late 2019, and the payment incentives that are now associated with items like mechanically altered diet and swallowing disorder which were just added to the payment system when they changed it two years ago.

So, there is an acknowledgment by the developer that this conversation ought to continue with a pretty sharp focus on at least those things and some of what we've discussed today.

Co-Chair James: So, what are our choices in voting? Do we let this pass for now with the understanding that in three years they'll come back with the risk adjustment in hand?

Co-Chair Yealy: Yes, we -- our decision is to vote up or down. And then, move from that with not anything contingent.

I wonder, John, if we should ask the developer if they have any additional focused replies. This is a common thread of concern. It has to do with subpopulations and adjustment so as not to overly penalize any particular group or location.

This might be a good chance, Theresa's just talked about some of these considerations that have already come into play since the last approval of the measure. I wonder if we bring the developer in for 60 seconds of a focused reply here?

Mr. Nagavarapu: Hi, everyone.

Ms. Lin: This is Cheng Lin.

Mr. Nagavarapu: Go ahead.

Ms. Lin: Yes, thank you for the opportunity.

First, we want to mention that risk adjustment is something not raised particularly in the 2015 maintenance. And, Acumen also monitors the help desk for the nursing home measures and we didn't receive any particular inquiry about risk adjustment.

However, in this maintenance cycle, this is something we paid special attention to, as you can see from the appendix, we have drawn three models from more simple ones to the complex ones. The most complex one, we included all active diagnosis on the MDS which includes 44 items.

And, we see that there -- of the three models, really, there's a small impact in terms of provider ranking for nearly all of them and the movement for 90 percent of providers, they stay in the same decile.

And, the model really pointed to that the MDS items of the highest risk is swallowing disorder and mechanically altered diets. But these were the two that are associated with some payment incentive. So, we were highly recommend that we hold on and further evaluate.

The other item we didn't incorporate into risk adjustment but we need to pay attention to is cancer. This one is a new item recently added. So, just to keep the COVID impact out of the measure evaluation, we don't have enough data to evaluate the impact of cancer. This is something we'll pay more attention to.

Regarding the emphasis of specialized facilities, we recognize even though the risk adjusted models show that over 90 percent of facilities stay in the same decile, and for risk adjustment focusing on cognitive status using dementia, depression, and Alzheimer's, 97 percent of providers stay in the same decile.

We acknowledge that the remaining three percent, there could be a concentration of specialized facilities.

But this is something implementation wise that may be difficult to adjust to risk adjustment.

We want to mention that this measure is totally reported through Nursing Home Compare and this is the site for provider -- for families and patients to select providers. So, if someone is looking for a specialized facility, this may be some consideration in that this kind of selection of provider perspective may not be something we want incorporated into risk adjustment.

Because that's just something we want to raise here and I'll turn it back to the group.

Mr. Nagavarapu: And, really quickly, this is Sri from the measure developer. I just want to add that Dr. Yealy, Dr. Schoenborn, and Dr. Yu from the committee made very strong arguments for reasons why conceptually we might really hesitate before risk adjusting for certain clinical factors.

If you think that certain actions are modifiable, and for that reason, we want to be extremely careful, that is, for risk adjusting for any of these factors.

So, while we'll keep an eye on this and continue to take into account your feedback, we'd also want to recognize the opinion of all the folks on the committee who have noted that we need to be careful not to create unintended consequences by risk adjusting and make sure that we're targeting these populations properly.

Co-Chair James: Are there any further comments before we go to vote?

Co-Chair Yealy: Yes, one comment in there that where Geeta notes that, you know, stratifying on diagnosis and comparing it might be of help. I wanted to get that into the record. That's exactly what we've been talking about with, you know, we can lump and dice.

My view of the conversation is it's really the same core concern. What about different people have different trajectories and impacts and how much would that influence anything drawn from the measure?

Ms. Bal: And, John and Don, if I can just jump in really quickly just to clarify that the vote that we're about to take is if we want to accept the SMP result of I believe moderate was their rating.

And, if we -- we must get greater than 60 percent saying, yes, they want to accept it. If we don't get the greater than 60 percent, we will do a normal vote, so the high, moderate, low, insufficient, and we'll go from there.

But, again, and I Just want to emphasize that we should be voting on the measure as is, not the measure that we're hoping it can be. But we can definitely keep track of the concerns and make sure they're highlighted in the technical report set for future iterations of the measure. We can make sure the developer knows to come back with, you know, having done some more evaluation of that area.

Co-Chair Yealy: Yes, and Poonam, help me just understand, Geeta brought this up, I think, it was consensus not reached, not moderate, is that correct?

Ms. Bal: I'm sorry, I'm getting confused. You're right, it was consensus not reached. We'll do the normal vote, apologize, we'll do high, moderate, low, and insufficient instead of accepting the SMP vote.

Co-Chair James: Okay, let's go ahead and do it.

Ms. Ingber: All right, I will open the vote. Okay, voting is now -- oh, actually, I want to double check if a committee member joined the call before we actually call the vote. I think that might have happened, Poonam.

Ms. Bal: Arpana, are you on the call? We see your name here, and we just want to make sure that you could introduce yourself and provide just any disclosures that you might have before we jump into the vote.

Member Mathur: Yes, I am on the call, thank you for having me. No disclosures at this time. And, I will be in the voting polling session also.

Ms. Bal: Perfect.

And then, Arpana, I don't know if you heard, we did send a new voting link out earlier on I think around the 10:30 --

Member Mathur: Yes, I have that.

Ms. Bal: Perfect.

Member Mathur: I do have that, thank you.

Ms. Bal: Thank you so much.

Member Mathur: All right.

Ms. Ingber: And, we're good to go. Thank you, everyone, apologies for not clarifying that first.

I see 17 results, I think we're waiting for 20. Okay, I'll just calculate the results.

Okay, voting is now closed on Measure 0689 for validity. We have one vote for high, 15 votes for moderate, 3 votes for low, and one vote for insufficient. Therefore, the measure passes on validity.

Thank you, everyone.

Co-Chair James: Okay, let's go on to feasibility.

Member Edelstein: Okay.

Co-Chair James: Does the discussant team have comments?

Member Edelstein: Well, I'll just say that the feasibility conversation should be much shorter and easier. This measure, as you've already heard is drawn from data provided already by long-term care facilities on the minimum data set. It is a mandatory form. No additional work required on the part of the facility that they aren't already doing.

Co-Chair James: Comments from the committee? I don't want to rush anybody. Are we ready to vote? Then let's do it.

Ms. Ingber: Okay, apologies, there's many clicks involved.

All right, voting is now open on feasibility for Measure 0689. Your options are high, moderate, low, and insufficient.

Just calculating. All right, voting is now closed on Measure 0689 for feasibility.

We have 15 votes for high, 5 votes for moderate, 0 votes for low, and 0 votes for insufficient. Therefore, the measure passes on feasibility.

Co-Chair James: We're on a roll. Let's go on to usability and use. Comments there from the discussant team?

Member Edelstein: Just to reiterate what the developer mentioned. The measure is already reported publically on Care Compare, formerly Nursing Home Compare. And, providers also do get preview reports through CASPER as well as the Provider Data Catalogue.

Co-Chair James: Questions from the committee? You guys must be getting hungry. Let's vote.

Ms. Ingber: I'll just clarify that we're voting on use first and then usability.

So --

Co-Chair James: Thank you.

Ms. Ingber: Yes, so, voting is now open for Measure 0689 on use. Your options are pass or no pass.

Just waiting on a few more. And, one more. I'll just pause for a couple more seconds. Oh, someone might have stepped away. Okay, so, we'll have 19 rather than 20. Great, Thank you. I will close the voting poll and share the results.

For Measure 0689 voting is now closed on use. We have 19 votes for pass, 0 votes for no pass. Therefore, the measure passes on use.

Co-Chair James: So, let's go on to usability.

Member Edelstein: I don't really have any additional comments to make. The weight loss measure is pretty consumer friendly, easy to understand.

Co-Chair James: Certainly. Are there comments or questions from the committee?

Seeing none, I would recommend we go ahead and vote.

Ms. Ingber: All right, voting is now open on Measure 0689 for usability. We have -- your options are high, moderate, low, or insufficient. And, again, we're looking for 19 votes.

Okay, I'll lock the poll and share the results. Okay, voting is now closed on Measure 0689 for usability.

We have 10 votes for high, 9 votes for moderate, 0 votes for low, and 0 votes for insufficient. Therefore, the measure passes on usability.

Co-Chair James: Okay, the last is suitability for our endorsement. Are there comments? Questions?

I don't see any, let's go ahead and vote then, please.

Ms. Ingber: Okay, voting is now open for Measure 0689 on overall suitability for endorsement. Your

options are yes or no.

Okay, I'll close the poll and share the results.

For 0689, overall suitability for endorsement, we have 18 votes for yes and 1 vote for no. Therefore, the measure is recommended for endorsement.

Co-Chair James: Okay, thank you, everybody. We walked through the swamp and came out dry, it's a good thing.

3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel (Centers for Disease Control & Prevention (CDC))

All right, let's -- baring anything else we need to do, let's go on to Measure 3636.

Ms. Buchanan: John, we're actually thinking that this might be a good time for a little pause, a ten minute break for everyone. How does that sound?

Co-Chair James: Sounds fine to me. Anybody else?

Ms. Buchanan: And then, I think --

Co-Chair James: We have a break scheduled at some point today?

Ms. Buchanan: Yes, we have a lunch break scheduled for 12:30 and so, I think that if we take a break now and then push lunch back, I believe, anyone on my team can jump in.

Co-Chair James: How far back are you going to push lunch?

Ms. Bal: It'll depend on the discussion, John. But, I think, you know, just with everything we've got, just a ten minute break would be good for everyone and then we'd be back at, well, about eight minutes and be back at 12:00 and hopefully we can conclude the discussion around the 1:00 or earlier, depending on how, you know, how much discussion we have on this

measure.

Co-Chair James: Okay, well, I'm of the mindset that probably committee members make for swifter deliberations. We'll see. Anyway, let's take a break. Okay? Ten minutes?

Ms. Bal: Yes, thank you. We'll see you at 12:00.

(Whereupon, the above-entitled matter went off the record at 11:52 a.m. and resumed at 12:01 p.m.)

Ms. Buchanan: So, now, I believe I'll turn it right away to you, John, to introduce the measure.

Co-Chair James: Okay, thank you, Erin.

So, this measure has to with vaccination rates, coverage among healthcare personnel and healthcare facilities.

It happens every quarter and there's some things that I think need clarifying in here but we'll work that out as we go along.

I'm going to bounce it right over to the discussants, but first, let me ask, are there any comments before we actually go to the discussants?

Ms. Buchanan: I think we have the developer do three to five minutes before we jump into the discussion.

Co-Chair James: Oh, yes, okay. Somehow I skipped that. Anyway, CDC on the line?

Dr. Budnitz: Yes, good afternoon. This is Dr. Dan Budnitz. Am I coming through clearly?

Co-Chair James: Yes.

Dr. Budnitz: Great.

I want to thank you for the opportunity to support discussion of this measure today.

I'd like to make three key points to introduce the measure. First, healthcare personnel, COVID-19 vaccination is a critically important patient safety issue right now and we believe likely will continue to be in the future.

Healthcare personnel COVID-19 vaccination is key for protecting healthcare workers from infection, thereby preserving an adequate healthcare workforce to care for nursing home residents and patients.

Healthcare personnel COVID-19 vaccination is also key for use in conjunction with other infection control measures to protect patients and nursing home residents from infection as well.

At the time the measure was originally submitted, there were no systematic reviews of the empirical evidence for COVID-19 vaccine effectiveness.

However, since the original submission of this measure, systematic reviews of the impaired data on effectiveness of COVID-19 vaccines have now been published. These all have concluded that the primary series of a COVID-19 vaccination is highly effective in preventing infection. The vaccine effectiveness estimates exceeding those of the influenza vaccine.

And, indeed, healthcare personnel influenza vaccine has been an NQF endorsed quality measure, Number 0431, in use for a decade.

So, based on three key systematic reviews, which now identified and summarized in public comments on the NQF website, the quantity of evidence of vaccine effectiveness is high with 17, 19, and 30 studies examined in each of the three reviews respectively, the quality of evidence is moderate, as these systematic reviews involve many studies which included estimates that were that were adjusted for potential confounders, and the consistency of evidence is high with all three systematic reviews concluding moderate to high vaccine effectiveness after a primary series.

We do acknowledge that limitation of these systematic reviews is that they do not provide formal grading of the evidence.

And, as I have mentioned, we have posted a summary of these three systematic reviews in public comments on the NQF website and we hope the committee will be able to take this into consideration that there are now systematic reviews demonstrating vaccine effectiveness when considering this measure.

There are also a handful of recent individual studies specifically on nursing home staff vaccination and new ecologic studies linking vaccination to reduced resident illness.

The second point I'd like to make is that this is not a wholly new measure with uncertain feasibility and usability. The data collection for this measure is feasible and widely used.

The measure of healthcare personnel COVID-19 vaccination was built by reusing data elements of the NQF endorsed Healthcare Personnel Influenza Vaccination measure which has been in use for over a decade in multiple types of facilities.

In addition, reporting COVID-19 vaccination coverage through the National Healthcare Safety Network has been part of the public health emergency response led by CMS and CDC for over a year.

Data are currently being reported by 15,000 nursing homes and thousands of other facilities.

And, of course, the healthcare personnel COVID-19 vaccination measure does have some modifications from the NQF endorsed healthcare personnel influenza vaccination measure, 0431. The most obvious being that these COVID-19 vaccinations is reported quarterly rather than annually.

Third, we all know that we remain in the midst of an

evolving pandemic which makes COVID related safety measures challenging to develop, but even more important, to adopt and implement for the safety of patients.

With new SARS-CoV-2 strains, there are recommendations for additional vaccine doses. Indeed, in the future, there likely will be new vaccines approved and new vaccine scheduling recommendations.

But new evidence for additional doses is not going to change the requirement for an initial primary vaccination series as a minimum baseline. Indeed, the primary vaccination series continues to be shown to be highly effective in preventing symptomatic and serious disease and deaths.

And so, we should not allow the project to be the enemy of the good, especially in the context of this pandemic. We can adopt this quarterly measure of healthcare personnel vaccination with a primary COVID-19 vaccine series and also plan for another measure or later revise this measure if or when there's sufficient supporting evidence to do so.

So, I thank you and I look forward to the discussion of this measure.

Co-Chair James: Okay, are there discussions or comments? How about if we go on to the lead discussants?

Member Thraen: All right, that would start with me. There's three of us, Curtis and Anne. We'll each take a section and present as we move through the criteria.

I'm starting with the overview of the evidence and importance in gap. And, first, I want to share with you, thank you for those comments from CDC. This is a CDC sponsored process measure.

And, it raised a couple of issues in our review just for

NQF sakes in terms of future planning. When there is a public health emergency like COVID and the evidence is not readily available quite yet, this process that we have currently is not particularly sufficient to manage that particular issue.

So, we're asking for consideration of some other process that maintains flexibility for short cycles for review during change of a public health emergency, possibly a PDSA cycle. So, that's just sort of an overall comment.

Specific to the measure, at the time that the evidence was presented and submitted for this measure for NQF consideration, the review team determined that there was insufficient evidence.

Now, you've heard from CDC's testimony, developer's testimony, that other evidence has been developed. They referenced three systematic reviews that are now available on their website. That was not included in this submission packet.

We found there was some confusion about the focus of this particular measure. The measure's entitled, All Healthcare Providers, but the evidence that's provided is really specific to long-term care. We're not clear whether or not the intention of this measure is for all healthcare providers regardless of facility or if it's really intended for the nursing home environment long-term care as referenced in his presentation.

So, clarification about the focus of the measure. Clarifying definitions on what constitutes a complete vaccination course. You used the term primary vaccination series. That was not found in the documentation for definition purposes.

It is a process measure, not an outcome measure. We do believe that this is a critical public health emergency measure, and therefore, has importance associated with it. But that the review of the packet itself did not provide sufficient evidence as also

determined by the NQF review process.

And then, finally, we did feel that there was some vagueness and there was a lot of reference to the ACP recommendations that was being presented as a substitute for evidence. And, the question of timing of this, again, the submission of the evidence or this packet versus what's occurred since that time and that the ACP recommendations are really what this measure is resting, and the packet is resting itself on.

And then, finally, we did see that there is an opportunity to improve. We did see varying rates in terms of performance gaps by type of provider, physicians, therapists, nurses' aides, ancillary service employees, et cetera. But we, again, we were unclear on whether or not you're talking strictly about nursing home long-term care or a broader spectrum in terms of who you're including in that effort.

I'll stop there, that's our overview, for any comments or questions.

Member Sood: So, I'm not sure, I did raise my hand.

I wanted to ask a question and comment, if that's okay. So, I've been thinking a lot about this evidence criteria for NQF and thinking through how that's probably one of the most important pieces that we do here in the committee.

Because the context matters. So, if you're weighing a low, you know, low importance metric like did you fill out the race in the health record, it's going to be very different than sepsis or are you trying to include vaccination in a more of a pandemic?

So, I think the evidence piece has to be thought through in context. It's not just demonstrating variability. It's not just demonstrating a before/after study or randomized control trial.

So, in that context, I guess I would say, I think the evidence is very strong. We know that vaccines are

in clinically effective, that that's a -- just an epidemiologic meteorologic kind of view, I think that's -- there's very strong evidence that vaccines are effective.

There's circumstantial evidence that if you vaccinate healthcare workers, that's going to reduce or be associated with less cases in the healthcare patients and in long-term care facilities. And, I see that the CDC put in the chat they also have done a study, observation, or looking at the evidence that this is going to have an impact.

So, I think methodologically there's strong evidence.

And, I guess the last thing that I want to add is, this is my playground. I am a hospital epidemiologist in a hospital for the Rhode Island Public Health Department as a consultant. The importance of being able to measure, incentivize, and influence vaccination in healthcare workers, in the middle of a pandemic, being able to gather even this much evidence I think is really quite remarkable.

And, I think that would be really -- I think, as a -- from a personal, professional point of view, I think this is a very important measure.

Member Collins: And, John, I'll jump in, too. I'm one of the discussants as well.

I agree with exactly what Dr. Sood said. Just to comment that in terms of perspective here, according to the CDC, over 3,500 healthcare workers have died from COVID.

And, if you put things in perspective, the risk of dying here is a real variable, but the last CDC estimate was vaccinations, just the primary series, you had a 15 times higher risk of dying if you were unvaccinated.

So, I think, you know, just kind of putting it in perspective here what we're talking about and what that means for healthcare workers and patients.

And then, and, you know, I guess, the next question is, Dan mentioned that there's new systemic reviews I think some of us were familiar with some of that data.

Kind of a question for the NQF staff, is, are we able to factor that in in terms of our discussion right now?

Ms. Bal: So, you are to vote on the measure as submitted. I think to keep this point, we offer the insufficient with exception option with the idea that based off of your expertise and knowledge of what's happening, even if there's not strong evidence or the evidence that was provided in the submission doesn't provide all the information that you need, then you can vote to move this forward with an exception.

So, while you should be reviewing it as is, the developer has provided that link in the chat. So, if you want to factor that into your decision, you can.

Member Thraen: Thank you for that because it's really my view, this is a timing problem, I think in all of our views.

Co-Chair James: And, to be clear, right, for the NQF staff, in terms of the algorithm and how you assess the evidence, it really comes down to the Algorithm 1 in our criteria document and was quite heavily based on the lack of a systemic review submitted at the time, is that correct, in terms of the flow chart?

Member Thraen: The developer indicated they would be happy to answer some questions. So, specifically around the issue of the intent of the measure for all healthcare workers versus long-term care. Could you comment on that?

Ms. Bal: Sorry, before, Dan, you go, so I will be Don for this since he had to step away. And so, in following the process that we've established with questions, we'll do that first and then, Dan, we'll jump to you to respond in just a moment. But I just want to make sure that we're following our -- the

process that we've outlined.

So, before we jump into that, I just want to clarify that, Geeta had asked a question about what the primary series was and what was up to date.

The developer did already respond and I think I know you all brought this up, the developer did respond that it's only that you had your initial set, so the two with Pfizer and Moderna and one with J&J and then if -- that does not include having a booster. So, just having that initial treatment.

And then, let's see, I just want to make sure there's no other questions in the chat that weren't answered.

All right, so we got several mentions of the report and so I think other than that, the only outstanding question is what, you know, which settings does this apply to? So, Dan, if you could respond to that?

Dr. Budnitz: Sure, happy to.

So, I think as was mentioned, we are a little bit limited by the NQF process. One of them being that the validation and reliability data must be in the setting for which the measure is indicated.

And so, we have reliability and validity testing data in the nursing home and long-term care setting. And so, that is the only data that is presented as part of the measure.

So, strictly speaking, the measure is for the long-term care or nursing home setting.

I very much appreciate the comments, Dr. Sood and Dr. Collins, providing us context that this measure exists in and do appreciate that context. Thank you for bringing it up.

I think in the chat, we did answer the question that the completed primary series, you know, is one adenovirus Jansen vaccine or two mRNA vaccines. And, I do want to emphasize that this can be the

measure right now. It doesn't preclude a second measure or updating the measure as an epidemic develops.

Thank you.

Member Thraen: There's another question in the chat about, is there evidence that a process measure like this will actually impact vaccination rates, for example, in your other vaccination rate -- vaccinations that you've already had in place for the last ten years?

Dr. Budnitz: So, the short answer is, yes. Influenza vaccination has been linked to reduced outbreaks in nursing homes. And, there is, again, most recent data, the most recent published that New England Journal article, I think it was just a week or a couple weeks ago that did associate decreased case rates when there's high community prevalence in facilities that had high vaccinate rates of healthcare personnel.

Member Thraen: Theresa, does that answer your question?

Member Edelstein: Somewhat. Mine is more a question of interpretation by the consumer about what it means about the facility when they're looking for a facility for their loved one.

You know, there are a lot of things that enter into a staff person's decision about accepting a primary series of vaccine and we see it all over the map. And, it also extends to boosters because in, at least in our state, boosters are now required.

So, I am just concerned that the consumer interpretation will equate low vaccinate rate with poor facility and that is a false equivalent.

Member Sood: If I could just add that maybe not be the facility's fault. But if it was my family member, as a quality metric, I absolutely want to know what the

facility vaccinate rate was for the staff and if I had an option of going to a facility that had higher rates of vaccination. I would absolutely be choosing that ages are the biggest risk for severe disease.

I do understand that it's not the facility's fault, that it leads to staffing changes and staffing concerns, et cetera.

Co-Chair James: Okay, Yanling?

Member Yu: Thank you, John.

I think the evidence is overwhelming in support to have a measure like that. And, you know, the vaccination have been shown to reduce series consequences, deaths, and hospitalizations.

So, I think, and also, you know, to fight the pandemic, we have to have the data. Without data, we don't know what's going on. And, I think even though, you know, the pandemic is like an emergency situation, but from the trajectory, there is, you know, evolving into like an epidemic like a flu. So, I think there is a valid approach to really continuing to collect the data and to see how we're doing to battle this coronavirus.

And, as a consumer, and like Geeta said, I definitely want to know where is the low vaccinate rate and where's the high vaccination rate. I do not want to -- my family members to get infected in a facility or long-term care that have a very low, you know, vaccination rate.

So, I think for consumers, this is a very valuable information that they can use to help them out to making informed care decisions if they have the choice.

Co-Chair James: Nancy, you have a hand up. Thank you, Yanling.

Member Schoenborn: Thank you.

I think it was really to clarify what maybe Iona asked on my behalf earlier. So, I have no doubt that, you know, higher vaccination rates are going to lead to less outbreaks but is there any evidence that reporting on the vaccination rates such as this measure proposes to do actually changes or improves the vaccination rates?

Member Thraen: I think that's aimed at the developer.

Dr. Budnitz: Yes, sure. So, thank you for that question.

I can -- I'd have to come back to you with that if anyone on the committee does know of that specifically reported. I don't think there's any systematic, you know, reviews of reporting leading to those changes. I think it's a pretty high bar. I mean, I don't know if every measure that's NQF endorsed has evidence that reporting that measure has changed vaccination rates especially new measures that are coming before.

Member Sood: No, I'm sorry, I'm going to jump in and not chat only because it's an answer to a question and not asking a question. And, I guess I would say to answer Nancy's question that we -- there are data in hospital settings that when you include influenza vaccination as part of the value based program, things that are publically reported, the compliance rate goes up dramatically. So, I would consider that as part of evidence that reporting it and including it in a value based purchasing program does impact.

Dr. Budnitz: So, I apologize, just to clarify, I'm sorry. I was asking -- answering the question, does it change reporting changing incidents of cases in patients. I don't know if we have data on that. Certainly reporting does change vaccination compliance rates by the healthcare personnel

themselves. I just wanted to make that clarification.

Co-Chair James: There may be an unintended consequence here, a major hospital system in my city decided that if you weren't vaccinated, you were going to lose your job. And so, they discharged, I forget what percentage of employees, but it was not insignificant that refused to get vaccinated.

Now, my concern is there's all this press about nurses being stressed to the hilt, especially. And, is there a chance that this requirement could force hospitals to fire those who are not vaccinated and thereby become understaffed?

Member Thraen: I think that addresses the usability and use of the measure component. I don't think that addresses the importance of the measure. So maybe, John, I think we might bring that up during usability and use.

Co-Chair James: Okay. I was just thinking it might be an unintended consequence and I'm not sure where that lies, Iona, thank you.

Member Thraen: Are there any other questions for importance of this measure? I think we've covered the issues that we saw.

Member Falvey: Iona, I do have one quick question. I'm sorry, I apologize. I'll keep it very brief.

It is an important issue, but the recent Supreme Court ruling about CMS facility that received money from CMS, you know, for Medicare or Medicaid services requiring healthcare personnel in those settings to be vaccinated. How does that, you know, how did the developers think that the evidence that they have is going to be affected by the fact that everybody's going to be required to have that kind of vaccination coverage or facilities that accept Medicare and Medicaid money?

Because I anticipate that there's very few nursing

homes that don't accept Medicare or Medicaid as a payment source. So, I'm just curious how that evidence they think lines up with that recent ruling?

Member Thraen: I think that's a developer response. We did have a brief discussion of that as reviewers and backed away from it because it wasn't contained in the packet. So, I'll let the developer respond.

Dr. Budnitz: So, seeing that this case went to the Supreme Court, it's probably a little bit above my pay grade. But I would just note that there still are challenges to vaccine requirements and I don't know if this is the last word that we will hear in that recent decision. So, I think that is yet to be determined if there indeed will be standing requirement for eternity or not.

Member Falvey: Thank you.

Member Thraen: Any other questions before we go to vote? Sorry, John, I stepped in.

Co-Chair James: That's okay, Iona.

Member Thraen: Old behaviors die hard.

Co-Chair James: No, I echo Iona's appeal. Are we ready to go vote on importance?

Okay, hearing no further comments or hands up, let's go do a vote.

Member Thraen: Point of clarification for NQF, do we have to vote -- are we supposed to vote on accepting the review? This is a process measure, not an outcome measure, so the Scientific Panel did not review this one, but the NQF staff did review this. So, do we -- what are we doing?

Ms. Buchanan: That's not until we get to scientific acceptability. Right now, we're just voting on importance to measure and report.

Member Thraen: Perfect, thank you.

Ms. Bal: And, just so everyone knows, staff ratings are always just suggestions. You'll never vote to accept our vote, you will just vote normally for any of the criterion.

Member Thraen: Thank you for that.

Ms. Bal: And then, just one more clarification. I know this comes up often, in order to vote for insufficient with exception, we need greater than 60 percent to vote for insufficient. If we -- if there's a combination of moderate or insufficient or a combination of low and insufficient, that would not result in an option to vote for an exception. So, I just want to make sure that's clear before we go in the vote. So, it's not a -- yes, you had a question?

Member Sood: Sorry, just to clarify on what you just said, because that makes it sound like it may be better to vote insufficient because then you can go for insufficient. But I think if I'm understanding correctly, if you don't vote for insufficient, if you vote for high or moderate, it will pass. If you vote for insufficient then there is that option as well. Am I understanding that right?

Ms. Bal: Yes, that's correct. If we have greater than 60 percent vote high or moderate, it will pass without an exception. And, if we have greater than 60 percent for insufficient, we have an option to vote with an exception. But any other combination of that would result in the measure going down.

Member Myrka: This is Anne, I just had one comment and I want to go back to NQF being responsive to public health emergencies.

You know, and I think that perhaps the insufficient with exception might be a good option for thinking about could we -- could this have come earlier to the committee review because we're in a public health emergency? Could we have taken it out of the normal course of review and reviewed it, say for example, within a month of submission?

You know, I'd like NQF to think about that. So, thinking about getting these PHE-type measures done more quickly even if there is insufficient evidence. I think this has a high face value anyway and you would have to, you know, we'd have to consider those options, too.

So, I just wanted to reiterate that for NQF. Thank you.

Ms. Bal: Yes, thank you for that feedback. We will definitely take into consideration as we think through potential improvements. But, unfortunately, at this point, we do have all measures follow the same process.

Co-Chair James: May I ask a question? What if the vote is split between high, moderate, and insufficient? Neither gets 60 percent?

Ms. Bal: So, if we have CNR, then the measure would move forward. Oh, I'm sorry, one more clarification, there is no high since there was no scientific review, there is a systematic review, the option of high is not there. So, it's only moderate, low, and insufficient.

And then, if we are not able to get greater than 60 percent either way, if it's between 60 and 40 percent, it would be considered consensus not reached. We would continue discussing the measure. We would not vote on overall suitability and then we could come back together during the post comment call to revote on evidence at that time, hopefully with more information but to support the committee's decision.

But those are the options, so we would either need greater than 60 percent moderate, greater than 60 percent on insufficient to consider the insufficient with exception and if low and insufficient are greater than 60 percent, the measure does not pass. And, if it's between 60 and 40, it is consensus not reached. Lots of nuances there, but we can, you know, always take the vote and then talk through what that means.

Co-Chair James: All right. Is everyone ready to vote?

Ms. Bal: Sorry, John, there was a comment that came into the chat, Curtis had a question about, can a developer summarize the systematic reviews recently provided?

Since we were not able to get those in earlier, we can't really base our decision off of those. However, since the developer provided it to you, you're free to use that as guidance for your decision.

And then, just a reminder, insufficient with exception is the same as passing the measure as moderate. The measure is sustained, but the measure is important and should be moved forward and so on.

Co-Chair James: Okay, does everyone understand the consequences of their vote? It's not straightforward.

All right, Hannah, are we ready to vote? One last call for any questions, ready to vote on importance?

Ms. Ingber: All right, thank you, everyone. Voting is now open for Measure 3636 on importance to measure and report evidence. Again, your options are moderate, low, or insufficient.

I believe we're looking for 19 votes.

Co-Chair James: Why am in not getting this?

Ms. Bal: You might need to refresh your screen. That's a common -- so a pretty common solution to the issue. But if you continue to have difficulty or anyone else does, please let us know.

Ms. Ingber: Okay, just waiting on one more I believe.

Ms. Bal: And, I think we're actually looking for 18, we can move on.

Ms. Ingber: Oh, okay.

Ms. Bal: We have two people missing.

Ms. Ingber: Apologies. We'll calculate the vote now.

All right, voting is now closed on Measure 3636 for importance to measure and report evidence.

We have 12 votes for moderate, 0 votes for low, and 6 votes for insufficient. Therefore, the measure passes on evidence.

Member Thraen: All right, the next one is gaps, performance gaps. And, as I mentioned earlier, the -- looking at CMS, looking across individual provider rates, we do see differences with physicians at a 75 percent; therapists, 69; ancillary service employees, 58; nurses, 56; and, aides at 45 percent, rounded.

Also, looking at certified long-term care facilities, using the NHSN source of information, the first three quarters of 2021, we saw a means of around 35 percent quarter one; 56 percent, quarter two; and 64 percent, quarter three. So, the rates were going up.

And then, stratified by the size of the facility in quarter one 2021, vaccination coverage between the 10th percentile, nursing homes, and 90th percentile with 67 percent in gaps. And, in the lowest 10 percent of nursing homes, improved vaccination coverage rates was up to 38 percent and the 90th percentile to 89 percent, leaving a gap of about 50 percent.

So, there is definitely performance gaps either by facility and/or by provider type but also across time. So, as we have moved forward in the COVID progression, we're actually you're seeing the rates go up, so this may be a time variable that's impacting these performance gaps.

That's it for me.

Co-Chair James: Thank you, Iona.

Member Thraen: Oh, one other thing, it was preliminarily rated high for opportunity for improvement.

Co-Chair James: Are there comments or questions from the committee? Okay.

Ms. Bal: And, I'm not seeing anything in the chat.

Co-Chair James: Okay, I think we're ready to vote then.

Ms. Ingber: All right, voting is now open for Measure 3636 on importance to measure for 1-B, performance gap. Your options are high, moderate, low, or insufficient. And, we're looking for 18 votes.

Just one more, give people a couple more seconds.

All right, voting is now closed on Measure 3636 for performance gap.

We have 11 votes for high, 6 votes for moderate, 1 vote for low, and 0 votes for insufficient. Therefore, the measure passes on performance gap.

Member Thraen: All right, Anne will now take over in the reliability and validity criteria.

Member Myrka: Great, thank you.

So, starting with the numerator, the numerator statement is that the numerator for this measure consists of the cumulative number of healthcare personnel and the denominator population who, one, who have received a complete vaccination course against COVID-19 administered at the healthcare facility; or, two, reported in writing in paper or electronic form or provided documentation that a complete vaccination course against COVID-19 was received elsewhere.

The denominator statement is that the target population is the number of healthcare personnel eligible to work in the healthcare facility for at least one day during the one week data collection reporting period, excluding persons with contraindications or exclusions to COVID-19 vaccination.

The quarterly reported measure includes at least one week of data collection a month for each of the three months in a quarter.

The denominators are reported by aggregating the following categories, one, employees, those are all people who receive a direct paycheck from the reporting facility, i.e., they're on the facility's payroll; two, licensed independent practitioners which would include physicians and DOs, advanced practice nurses, and physician assistants who only who are affiliated with the reporting facility who do not receive a direct paycheck from the reporting facility; three, adult students or trainees and volunteers, include all students and trainees and volunteers aged 18 or over who do not receive a direct paycheck from the reporting facility; and, four, other contract personnel.

Facilities may also report on individuals who are contract personnel; however, reporting for this category is optional. Contract personnel are defined as persons providing care, treatment, or services at the facility through contract who do not fall into any of the above mentioned denominator categories.

Denominator exclusions are individuals with contraindications to COVID-19 vaccinations and individuals for whom the COVID-19 vaccine is not authorized or recommended.

The data source is the patient or encounter and is paper or electronic and that data source will vary by the entity of whoever is providing the vaccination and reported at the accountable entity level to the National Healthcare Safety Network or NHSN using a standardized form.

And then, NQF staff reported no concerns regarding specifications, definitions, or coding. So, I don't know if there's any questions about the numerator or denominator, specification statements.

Okay, I'll move on to liability. NQF staff evaluated this

process measure. The numerator reliability was tested during a three week period between 12/8/2020 and January 17th, 2021 at the accountable entity level using signal to noise analysis.

And, testing was performed in 869 nursing homes that reported vaccination to NHSN as well as reporting through the pharmacy partnership for long-term program, which you might all remember was the pharmacies that went into nursing homes and performed the vaccinations for the rollout in the beginning of 2020.

Pearson's correlation coefficient was 0.846 which was high and reported as linear and the Kendall's tau correlation coefficient was also report as high at 0.751. Both continued high after stratification by facility size, number of healthcare personnel, vaccination coverage, and reporting week.

There was a one to one linear relationship between the number of vaccinations of the two data sets and the difference between the two was not significant.

With stratifying by vaccination coverage rate, the range of mean differences were larger by tertile. So, a negative 4.81 vaccinations in the lower tertile and 3.57 in the higher tertile.

And, the differences noted between tertiles were likely due to NHSN receiving vaccination data from elsewhere and the PPP program only reporting entity delivered vaccinations. PPP may also have administered vaccinations to others that did not meet the healthcare personnel definition to ensure that all vaccines remained used so that there were none left over.

The Bland-Altman plot assessment showed no pattern of variability and minimal bias. And, no denominator testing was done as the same healthcare personnel denominator as NQF Number 0431, which is the influenza vaccine coverage among

healthcare personnel which also resulted in high reliability testing.

Based on this assessment, there is moderate certainty that this data is reliable by the developers and staff.

Any questions? Okay. Next --

Co-Chair James: I see Jason has his hand up.

Member Myrka: Okay.

Member Falvey: Yes, I mean, I just had a clarifying question. The contract personnel exclusion, I don't know, is that, and maybe the developer can clarify this, is that consistent across every reporting period for the same facility?

Because there are some staff groupings like rehabilitation staff, 70 percent of rehabilitation staff in nursing facilities are contract staff nationally, as most rehab is provided in that manner. So, that would mean that 70 percent of physical and occupational therapists in these facilities aren't captured routinely?

That's a lot of variability if facilities are reporting them if they're good and not reporting them if they're bad, it's -- it makes me wonder about the reliability and if they were reporting the same, you know, groupings, if it's just in house staff or contract staff or if they're tracking which ones are getting reported.

Nursing and CNAs aren't contracted at nearly as high a rate, but rehab and all of the related rehab groupings certainly are.

Dr. Budnitz: This is Dr. Budnitz from CDC. Would you like me to respond now to that?

Co-Chair James: Yes, please. I think that would be appropriate.

Dr. Budnitz: Yes, so, thank you for the question and

I think you identified an important point about contract personnel.

It's important to note that the independent practitioners like physicians, nurse practitioners, PAs, are included. We use the same categorization as what is used in the healthcare influenza healthcare personnel NQF approved measure.

So, the inclusion -- so, it is not required to report contract personnel. That being said, and again, we did this for approval to be as close as possible and consistent with what is being done for flu right now.

That being said, in the future, it may be possible to include contract personnel as when the flu vaccination measure was endorsed, it was ten years ago, there was fewer contracted personnel at that time. But that is something that could be considered in the future. But for this measure right now, we're being consistent with what is being done for influenza vaccination measure and they're not mandatorily included.

Member Falvey: All right, thank you for clarifying that and I am just of the same opinion not to let perfection to be the enemy of good but it is an important point of clarification I think with such a high number of contract staff in a lot of facilities.

Dr. Budnitz: Agreed.

Co-Chair James: Are there other comments or questions?

Member Charbonneau: John, I have a quick comment as a point of clarification that in inpatient rehabilitation facilities, IRFs, which are considered hospitals, that is not the case. So, physical, occupational, and speech therapists are not contracted in IRFs which are a higher level of rehabilitation than what you'd find in skilled nursing facilities. So, I just wanted to clarify that.

Co-Chair James: Thank you.

Other comments? Yes?

Member Edelstein: I would just like to acknowledge that the -- in the nursing and CNA area, largely due to staffing shortages and departures of staff from facilities in long-term care, there is now a higher rate of contract staff among nursing than ever before.

Member Charbonneau: We find the same in the IRF setting due to the pandemic as well.

Co-Chair James: Okay, other comments, other worries? Okay, having seen none, I guess the chat box is clean?

Ms. Bal: So, Geeta has a question about feasibility, but we'll hold that until that discussion point.

Co-Chair James: Okay. So, let's go on to vote on scientific acceptability, if we may.

Ms. Ingber: Okay, voting is now open on -- for Measure 3636 on scientific acceptability, reliability specifically. Your options are moderate, low, and insufficient because only data element testing was presented.

Just waiting for one more person.

Member Thraen: And, I think we lost Anne, she said she was having issues.

Ms. Ingber: Oh, okay.

Member Myrka: No, I'm back. I was able to quickly come back in.

Ms. Ingber: Okay, never mind, we are waiting for 18, thank you.

There we go, okay. I'll calculate the results.

Okay, voting is now closed on Measure 3636 for scientific acceptability, reliability.

We have 15 votes for moderate, 2 votes for low, and 1 vote for insufficient. Therefore, the measure passes on reliability.

Member Thraen: So, Anne's going to continue with validity.

Co-Chair James: Yes, Anne, go ahead and put up validity, please.

Ms. Bal: I think she's on mute.

Member Myrka: I'm talking away.

Regarding threats to measure validity, measure exclusions which do not include vaccine hesitancy were explored and there were no concerns.

Mean percentage of healthcare personnel with vaccine contraindications was 0.6 percent with the standard deviation of 2 percent. There were no concerns regarding meaningful differences in performance and mathematical modeling showed that increased vaccination rates would result in fewer symptomatic cases with facilities.

There were no concerns regarding compatibility of results as all data is entered using a standardized form into NHSN which is the ultimate data source for the measure.

Missing data is not of concern as greater than 98 percent of facilities reported all three months.

There is no risk adjustment.

Regarding testing, the testing level is a measure score using empirical validity testing correlating the measure scores with NQF Number 0431, as mentioned before, regarding influenza vaccination coverage among healthcare personnel.

And, the Pearson correlation coefficients in 1,807 nursing homes in quarter three of 2021 and comparing mean scores in 95 percent confidence

intervals for each 2021 quarter to each other and then to the scores for the 25th and 75th percentiles was performed.

The Pearson correlation coefficient was 0.4169 which shows a medium correlation and the results show a medium correlation by quartile.

Differences may include healthcare personnel declining of one vaccine and acceptance of another and community rates of COVID-19 may vary from influenza rates and impact healthcare provider personnel decision making.

In determining if there is statistically significant differences in performance measures, measure scores among measured entities, the developer calculated the mean quarterly COVID-19 vaccination coverage rate and inter-quartile ranges for the first three quarters of 2021 and the mean quarterly COVID-19 vaccination coverage metric was higher each of the first three quarters of 2021 based on non-overlapping 95 percent confidence intervals showing the measure can identify differences in performance across quarters.

Methods were appropriate for assessing the relationships and overall rating of validity was considered high.

Any questions? And, Geeta, I'm not a tech superstar, I'm just used to having a jump from teams to WebEx to Zoom and constantly do some saving maneuver to get back online, as we all are probably now.

Co-Chair James: Are we now ready to vote on validity? Thank you for the summary, by the way, Anne. That was thorough.

Member Myrka: Thank you.

Co-Chair James: Okay, let's go ahead and vote then, Hannah.

MS INGBER: All right, voting is now open for Measure

3636 on validity. Your options are high, moderate, low, or insufficient.

Just waiting on one more.

Is anyone having trouble voting or if other NQF staff could clarify someone? Oh, there we go, thank you.

All right, voting is now closed on Measure 3636 for scientific acceptability, validity.

We have 8 votes for high, 10 votes for moderate, 0 votes for low, and 0 votes for insufficient. Therefore, the measure passes on validity.

Member Thraen: Thank you, Anne, for that. And now, Curtis is going to take us out with feasibility, usability, use, and competing measures.

Member Collins: Yes, thanks, Iona.

So, feasibility, this measure is -- the data collection is in paper or electronic sources. It is already being performed and reported for a large number of long-term care facilities.

There are no fees and the reporting is familiar to most.

The preliminary rating for feasibility was moderate by the NQF staff. The other comments were generally agreeable. However, some committee members noted concerns with manual abstraction, exemptions, and their frequency of data collection and reporting.

Turn it over to John for discussion.

Co-Chair James: Yes, discussion on this point of feasibility? That was succinct, thank you, Curtis.

Ms. Bal: Oh, sorry, John, Geeta has a comment.

Member Sood: Sorry, I'm sorry to ask such I guess a redundant question, but really to ask the developers, because I think using the fact that it was feasible in

the middle of a pandemic when everybody had to do so much reporting on COVID might be a little bit different so that feasibility when we're coming to -- off of a surge and when it doesn't become such a priority.

So, if -- would it be okay, John, if we asked the developers to comment a little bit more about feasibility because I could imagine that it might be, I would love to hear their thoughts.

Co-Chair James: Okay, Geeta, did you ask for the developer to comment? Okay, that's what I thought. I have a little trouble understanding you.

Okay, if the developer would comment on feasibility, we'd appreciate it.

Dr. Budnitz: Sure, thank you very much for that question.

And, that is something that we really did consider carefully. And, in fact, there's reporting going on in nursing homes right now weekly. And, we actually didn't think that was feasible beyond, you know, this beginning of a pandemic.

On the other hand, the flu vaccination, the NQF endorsed flu vaccination measure is a yearly measure. And, we thought that was far too long for the results to have impact on, you know, vaccinations and to be available for consumers and patients and to use.

So, we kind of split the baby so to speak by doing this measure as a quarterly measure. And, that does not preclude the development of another measure in the future or modification of this measure to align with the influenza measure down the road once, hopefully, this pandemic's under control.

So, we really do appreciate it. We did have the sentiment and we tried to hew a middle ground.

Co-Chair James: Thank you. Any other comments or

questions?

Okay, let's go ahead and vote then. I don't want to rush people, but we've got to keep moving.

All right, Hannah, if you'll put up the feasibility vote, please?

Ms. Bal: Hannah, if you introduced the vote, we didn't hear you. You might be on mute.

Ms. Ingber: Oh my gosh, I'm so sorry.

Measure 3636 feasibility voting is now open and your options are high, moderate, low, or insufficient. And, we were waiting for 17 votes was what I said, apologies. But we have those votes in so I will share the results.

We have 10 votes for high, 7 votes for moderate, 0 votes for low, and 0 votes for insufficient. Therefore, the measure passes on feasibility.

Co-Chair James: Okay, Curtis, I guess you have the floor.

Member Collins: Yes, yes, moving on to use.

The developer indicates that this measure is currently in public use, public reporting, and regulatory and accreditation programs.

The long-term care facility data on the CDC website is trended over time and searchable by state. I would encourage people to check it out.

The preliminary committee rating for use was pass and the committee feedback largely agreed with this and had no concerns.

Co-Chair James: Okay, so does that cover use and usability?

Member Collins: I believe, and NQF can correct me if I'm wrong, but we vote on use and then usability.

Co-Chair James: Yes.

Member Collins: So, that was use.

Co-Chair James: Thank you.

Okay, so, questions about use? Okay, then, let's go ahead and vote.

Ms. Ingber: All right, voting is now open for Measure 3636 on use. Your options are pass or no pass. Another committee member has joined us so we're back up to 18 as our denominator.

Just waiting on two more. Still waiting on two more votes. We do have quorum and maybe someone is experiencing some technical difficulties, so I will close the poll and share the results.

Okay, voting is now closed on Measure 3636 for use.

We have 16 votes for pass and 0 votes for no pass. Therefore, the measure passes on use.

Member Collins: All right, and finally on to usability.

The developer states the vaccination rates amongst healthcare practitioners working in CMS certified skilled nursing facilities increased each quarter of 2021 as well as the number of facilities reporting also increased following the CMS requirement of May 2012.

No potential harms were identified by the developer. The preliminary rating by NQF staff for usability was high, and feedback from our committee largely agreed.

Open it up for discussions.

Co-Chair James: I don't see any hands up. Last call. Let's go to vote then.

Ms. Ingber: Okay, voting is now open for Measure 3636 on usability. Your options are high, moderate, low, or insufficient.

Okay, I'll give it just ten more seconds and then just in case someone's having any technical difficulties.

Okay, voting is now closed on Measure 3636 for usability.

We have 8 votes for high, 8 votes for moderate, 0 votes for low, and 1 vote for insufficient. Therefore, the measure passes on usability.

Co-Chair James: Okay, let's go on for -- to overall suitability for endorsement. Does the team wish to address that or Curtis, are you going to address that?

Member Collins: I wasn't planning to, but I would think from everything that we've heard today and how we voted that I would think that this will pass the overall suitability for endorsement and I would agree personally.

Co-Chair James: Yes, I don't see any particular variable to that or barriers. So, let's go ahead and vote on this if there are no hands up.

Ms. Ingber: All right, voting is now open for Measure 3636 on overall suitability for endorsement. Your options are yes or no. Again, we need a minimum of 16 votes but I'll pause just as we're waiting for 17.

All right, voting is now closed on Measure 3636 overall suitability for endorsement.

We have 16 votes for yes, 1 vote for no. And, therefore, the measure is recommended for endorsement.

Co-Chair James: It's up to the NQF staff, but I think we're probably about ready for lunch, aren't we?

Member Thraen: Did we address competing measures?

Co-Chair James: Oh, I'm sorry.

Ms. Buchanan: That will be later. Yes, that will be

after we discuss and vote on all of the measures.

So, now we're looking at a 30 minute lunch break. So, if everyone could be back at 1:40 Eastern Time, that will be great. Let us know if you have any questions in the meantime.

Co-Chair James: Okay, thank you, everybody and thank you to the developer for great discussions, appreciate that.

(Whereupon, the above-entitled matter went off the record at 1:10 p.m. and resumed at 1:41 p.m.)

Co-Chair Yealy: It's 1:41 p.m., it's probably time to get started back up again. We have everybody back.

Ms. Buchanan: Yes, welcome back to us, if everyone could turn their cameras back on? I'll close that. Also, if you're not speaking please mute yourself.

I hope everyone had a good lunch or a good break, depending on what you decided to do. Before we start off, I wanted to go into how we're going to conduct the discussion and the voting for the three measures we have this afternoon.

So, after our Co-Chairs introduced the measure and the developer provides the brief overview of the measure, we'll hand things over to the lead discussants.

And the thing about these three measures, as you all know, is they have a lot of similarities.

And in an effort to help the Committee realize those similarities and differences, and vote consistently across the measures, we've divided our lead discussants across the criteria as opposed to how we normally do where one lead discussant will follow through the entire measure.

So, each measure will be considered and voted on independently but the same discussants will lead the Committee through evidence and performance gap,

and the other criteria.

I'm not phrasing it properly, sorry. So, the same group of discussants will be going through a criteria per measure.

So, a group will be going through evidence in GAP and another group will go through scientific acceptability.

And then a third group will lead feasibility, usability, and use for all measures. So, as another reminder, similar to this morning, during the discussion, any questions from the Standing Committee for the developers will be noted by both Staff and the Co-Chairs, and will be discussed by the Committee first before going to the developer.

And then the Standing Committee will continue any other discussion that needs to happen.

If anytime during the discussion the developer team would like to address any inconsistent they hear or clear up some confusion, we ask that you please use the raised-hand feature and put a request in the chat.

The Co-Chairs will call on the developers to speak at the appropriate time. The full Committee will be discussing and voting on each of the measure criteria for each measure. And I'll pause there first and ask if anyone has any questions about that process? I think our discussants take up a pretty big portion of the Committee so I hope everyone understands.

I'm not seeing any, so, Don, I'll hand it over to you to introduce the measure.

Co-Chair Yealy: Thank you very much. And we're doing this in a newer approach, recognizing the overlap and trying to make sure that we approach this with as much rigor as each individual measure needs without any unneeded repetition, and there really will be.

I expect that the first measure will have the most

conversation and as we go through the conversation will become more and more focused. And there's nothing wrong with that as long as we're accomplishing what we want.

And at the end of everything we look at competing measures, of which these are being considered so there's nothing to compete with them yet.

We may have a different decision at the end of all the voting on all of that. So, I appreciate it and I'm looking forward to seeing this new novel method.

The first measure we're going to look at is really the same order that NQF created for us.

You all realize we could have picked any particular order of these because there's a natural flow, depending on whether you go from individual all the way through facility.

3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level) (Alara Imaging/University of California, San Francisco (UCSF))

We'll start with 3633E, which is radiation dose or inadequate image quality for a diagnostic CT scan in adults, measured at the clinician level.

The measure steward and developer was Alara Imaging and UCSF. This is a new measure, as are the accompanying ones with this. I'll turn things over to the developer to walk us through what they've created.

Dr. Smith-Bindman: Thank you. I was planning to introduce all three measures together at the beginning because the only difference is the level of attribution at the clinician group or facility.

I'm hoping I can take a few extra minutes of time to do so.

Co-Chair Yealy: We're considering sequentially. I expect you, like the discussants, will have a shorter version in the second and third presentations.

Dr. Smith-Bindman: I was going to have no version, I was just going to do one introduction if that works?

Co-Chair Yealy: I expect you will have a shorter version in the second and third presentations.

Ms. Bal: Don. Why don't we go ahead and let the developer do all three now? That way, we don't have go back later. It seems like the comments will be very similar?

Will that be okay, don?

Co-Chair Yealy: Okay.

Dr. Smith-Bindman: Thank you, my name is Rebecca Smith-Bindman, I'm a radiologist and epidemiologist and have been a faculty member at UCSF for 25 years.

For the last 15 years, my primary research area has been quantifying the radiation doses used for CT scanning and identifying ways to appropriately reduce excessive doses.

I identified the importance of quality measure as a way to advance the field over 10 years ago when I wrote 2 NQF-endorsed CT radiation dose measures.

The pediatric measure is still in use today and the Leapfrog reports on this measure for over 1400 hospitals annually.

But for the last 10 years, my research team and I have created a large CT radiation dose registry of more than 8 million exams collected from over 160 facilities, which has allowed us to quantify the radiation dose to understand the cause of the variation and to develop and study interventions to help facilities appropriately lower those doses.

The development of this quality measure was a natural extension of this work and the registry has allowed us to fully develop and test the adult measures that you're considering today.

In large part based on our 15-year track record in this area, under the Medicare Access and CHIP Reauthorization Act, MACRA, CMS awarded my team a cooperative agreement to develop these CT measures.

And please know I may speak of the three measures in the singular because they are identical except for reporting level.

The measure provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, which is a risk factor for cancer, while preserving image quality, and thus, the diagnostic usefulness of those exams.

The measure is expressed as a percent of CT exams that are out of range based on either excessive radiation dose or inadequate image quality relative to the evidence-based threshold based on the clinical indication for the exam.

This is an important point. Each exam is stratified and assess based on the reason for imaging, which is determined from the diagnostic and billing codes associated with the exam, water and bell.

This ensures patients receive the right radiation dose for their individual condition. The measure is primarily focused on radiation dose and was not intended to maximize image quality.

The image quality component was included as a balancing factor to ensure a minimum floor to quality. Also, the measure is an ECQM that uses automation electronic data extraction to minimize reporting burden.

In order to develop the measure, we gauged a very diverse and expert stakeholder group for our technical expert panel, including leadership for the American College of Radiology, the President and Chief Medical Officer for Radiology Partners, one of the largest radiology groups in the nation, representation from the Leapfrog Group, the Director of Quality Management at the Joint Commission, the Chief Medical and Scientific Officer for the American Cancer Society, the Executive Vice President at United Health, and a medical physicist who previously served as President of the American Association of Physicists and Medicine, and the Society of Imaging and Informatics in Medicine, and serves as Governor on the American the Board of Radiology.

The panel also included representatives from other medical specialties such as urology, emergency medicine and cardiologists, additional radiologicals and many individuals with methodologic expertise in measure development, electronic health data extraction and reporting, as well as several patient advocates.

We approach the measure development systematically and set forth several principles that we believe were important to create a state-of-the-art measure that I think highlights how our measure is broader than the NQF-endorsed American College of Radiology CT measure.

The four principles include the following key components. First, that the measure covers all body regions and all types of diagnostic CT scans, including the higher dose multiple-phase exams.

Second, while the primary focus was on radiation dose, that also included some mechanism to ensure radiation doses were not reduced too low so as to undermine image quality.

Third, the measure would cover two key process of care components that determine the radiation doses that are used. These include, A, the choice of imaging

protocol or in layman's terms the type of CT exam.

For example, whether a patient with a suspected pulmonary embolism is imaged with a single or double-faced CT. This decision is typically made by the performing radiologist.

And B, decisions regarding the technical settings used for that type of CT exam, which are usually at the discretion of the technologists or medical physicists who oversee and operate the machines.

As both of these components contribute to radiation dose, we believe a comprehensive quality measure must encompass both of these decision-making processes.

This brings us to the fourth and final priority in developing the measure. We wanted to make sure that the individuals and organizations who perform CT are aligned in efforts to improve quality and safety.

In the hospital setting it is the hospitals that typically own the machines, employ the technologists, and control the imaging data. The radiologists are typically part of a provider group which is separate from the hospital and they decide what protocols to use and supervise the technologists.

We realized early in the development process that challenges would occur unless these two stakeholder groups were aligned for reporting. And more importantly, for driving quality improvement.

With strong CMS encouragement, we created measures for both hospitals and physicians in order to position both measures for success. Lastly, I want to help you understand the quality gap that led to our creating the measure.

Over 91 million CT exams were performed in the U.S. in 2019, meaning the equivalent of 1 in 4 Americans is imaged with CT each year. This measure will

impact a significant portion of the U.S. population.

We developed the measure because the observed and extremely large variation in the radiation doses that are used for CT scanning. The doses that are used for particular type of scans, for example, a head CT to evaluate a patient after trauma or an abdominal CT from right-lower quadrant pain can vary by orders of magnitude depending on where the patient goes for the scan.

The variation is not driven by patient factors such as how large the patient is or why the scan is done, nor is it driven by the age or type of machine but rather by local decisions that are made of that exam type and machine settings.

I want to make clear these factors such as patient size and indication for scanning machines do influence the doses but they don't drive the dose variation.

The variation remains after accounting for these factors.

I also want to emphasize that appropriate dose scans can be obtained on all patients, all machine types, and more importantly, which my team has demonstrated in an NIH-funded randomized control trial that providing audit feedback to imaging facilities about their doses can achieve meaningful and sustained dose reduction.

These measures are intended to provide actionable feedback similar to what was provided in the trial.

I am joined today by several members of the development team who are available to answer questions, along with myself, including Patrick Romano from UC Davis who served as a close consultant, Yifei Wang, a biostatistician on my team, and Nathan Mazonson, CEO and Simon Rascovsky, Chief Technology Officer for ALARA Imaging.

Thank you very much.

Co-Chair Yealy: Thank you very much for the setup, I appreciate it.

We'll start with 3633E and the first portion of our duties is to look at the evaluation of importance to measure and report with this particular proposed measure.

We have a collection of four who are looking at it not only for this but for the other two measures, that include Joel, Nancy, Geeta, and Yanling.

I'd like to lead now and let us know what your summary is of the evidence and the performance gap.

Any of the survey comments from the Standing Committee or any other issues that you've uncovered that you want to share with our Committee.

Member Sood: Thank you, I'm going to start, I have great co-discussants with Nancy, Yanling, and Joel. I'm basically just setting the stage and the context for this metric.

And Joel and I are going to talk a little bit about what we learned in engaging our radiologists in our health system to again provide a little bit more of the background or context.

I know that unintended consequences really belong in the validity section but I think there's a component of that that also belongs in the evidence section. So, there will be a little bit of a focus on that.

So, just to clarify, thank you for that lovely discussion but I wanted to just point out that in speaking to radiologists at Johns Hopkins, and I think Joel spoke to his radiologists at Sentara, we learned a few things about some of the determinants of what goes into deciding the radiation dose.

Some of them were already mentioned but what I

learned specifically was that the higher radiation doses tend to be more multi-phasic studies and tend to produce greater resolution.

So, for studies that are looking for cancer in the pancreas or genital urinary system, the higher doses is often used in that scenario. For the experience of our radiologist, what I learned was that he had actually trained at another institution and was a little bit surprised by the amount of multi-phasic studies and the higher radiation studies that we were doing here.

But he has actually come to realize and appreciate that it's had a really terrific impact on reducing necessary repeat CT scans because of poor resolution and better delineation of tumors and other things related to outcomes and interventions that are being proposed.

I wanted to bring that up, and Joel, I think you had some additional information.

Member Bundy: Just to add to that, the radiologists here in my healthcare system were asking about the measure because they actually are already participating in the ACR registry, which has specifically looked at what doses should be for different types of CT scans.

And they're actually looking at software so a lot of the solutions were already there they were look at. So, they weren't sure exactly why to reinvent the wheel.

Of course, I'm confined with a healthcare system of 12 hospitals and it may be than free-standing radiology units. But that was sort of one of the questions about why are we looking at this anyway?

I think the rest of the discussants will talk more but we were talking about the evidence and the evidence was primarily based on pediatric cases.

However, this is really about adult imaging and so there was that dissonance there and maybe the developers could talk more to that later but why are we leaning so heavily on pediatric evidence when we're looking at really adult cases?

Member Sood: Joel, Nancy was going to talk about the evidence specifically but she's unable to make it.

Member Schoenborn: I'm still here for one minute. I can say something quickly and then ask Joel to fill in. Sorry, I have to step away for half of an hour.

Following on Joel's lead, just to look over the evidence that was submitted to support this measure, which is a process measure, and we're really trying to decide, I think the question for the Committee is whether there is sufficient evidence that links this intermediate process outcome, i.e. radiation exposure to an outcome, i.e. cancer.

And I would say the evidence that submitted, like Joel mentioned, are primarily in pediatric literature.

So, there are two systematic reviews that are submitted, one is entirely based on pediatric radiation exposure.

The other one combined a mixture but actually, when I pulled up the systematic review I would say most of them are environmental or work-related radiation exposures, workers around Chernobyl, for example.

So, there were only four studies that looked at medical radiation exposure and three of those four studies are again pediatric. So, this really just left one city that looked at adult medical radiation exposure.

So, I think that led to some hesitancy in answering yes to that question that's posed to the Committee.

So, even though I think intuitively there's face validity that extra radiation is likely to be harmful and if we can reduction unnecessary radiation that should be a good thing.

I don't really see in the evidence that is presented that radiation through CT scans in adults will lead to cancer. I'm sorry I have to jump off. I will be back at 2:30 p.m.

Member Sood: Just to piggyback on what Nancy said, we had talking points between us.

And Yanling, please chime in as you have other comments to add, part of what we were talking about in terms of the evidence is both the potential benefit, which as we talked about was more theoretical related to adult radiation doses and cancer, but also the unintended consequence piece.

I think this hasn't really been closely described in the measure although it sounds like Dr. Smith-Bindman did briefly talk about it.

But the unintended consequences of having to require a repeat scan, a delayed cancer diagnosis because the imaging quality was not adequate, or poor procedural outcomes because the scan obtained for an intervention was not good enough for what was being planned.

So, I think those aspects would also need to be considered. Joel and Yanling, other thoughts?

Member Yu: I appreciate this is a small group discussion prior to the presentation. I understand the unintended consequences, certainly a cancer you need sometimes higher doses in order to identify the tumor at the boundary and all that.

But I also think about it from a consumer point of view that is like everything else, it is risk versus benefit.

And we all know that radiation, it is well documented for children and there may be more risky and more vulnerable than adults.

But like Nancy said, at face value, the evidence shows that excess radiation does cause harm and no data

doesn't mean no data exists or no evidence exists.

I think maybe it's a matter of it to have measure-collecting data. We're doing the measure, we have the measure for children, for pediatric patient populations and it will be good actually to have a measure for the adult population.

And I read the public comments and I read from -- I forgot, was it ACR? The organization that raised a number of concerns but I also read about patients concerns about the access to radiation and the support to the evidence.

So, I would like to hear more from the sponsor about how you reconcile the evidence that your review provided evidence mostly focused on pediatric patients, and how you extrapolated that this evidence can apply to adults, and in terms of the risk and necessity for such a measure.

So, that's just my comments.

Co-Chair Yealy: One of our challenges always in doing the measure reviews is staying focused on the step-by-step process here. And I want to make sure that we go through the process the way that's intended.

So, right now it's the evidence and performance gap that we have to focus on first, not so much some of the other issues. And I realize how these often overlap and it's often very, very difficult.

Unintended consequences is important but it's just not where we're at right now with our conversations.

So, we are focused on is there evidence, I think as Nancy said, that this particular measure as it's constructed, links to an outcome if it's not directly measuring the outcome and that's there's a gap or a performance opportunity there.

Geeta, since you're taking the lead on this, is there anything specifically collective that you want to have

the developer answer to help move this conversation along?

Member Sood: Thank you for that terrific summary.

I guess the only question would be I see some links in the chat but if the developer could in just a very short sentence or two mention if there is good evidence that excessive CT scan doses is linked to adverse outcomes in older adults.

And I think we forgot to mention this specifically but there is definitely, as Dr. Smith-Bindman mentioned, variability in performance.

Co-Chair Yealy: The developer, that's a direct question now that you can engage on. Rebecca, I think you're muted.

Dr. Smith-Bindman: The requirement for the NQF application is to cite systematic reviews and the systematic reviews have focused on children but indeed there are several papers that have described cancer risks associated with CT scanning in adults.

A very important one published several years ago in Cancer Spectrum, a very high-impact journal, I put the title, shows that patients who have an increased exposure to CT scans have an increased risk of cancer.

But separate from that paper, I would disagree that there's not evidence from a large number of sources that patients exposed to the same radiation doses that are used in medical imaging are at an increased risk of cancer.

Amy Berrington de Gonzalez who leads the National Cancer Institute, Radiation Epidemiology Branch did an outstanding summary of the evidence across all imaging modalities, and estimated in a separate paper that the ballpark of three to five percent of cancer in the U.S. comes from medical imaging.

So, I disagree that there's not evidence. There's a lot

of evidence, there just happened to be systematic reviews in pediatrics. So, I think there's a lot of harm and there were a lot of points raised about the need for multiple-phase scanning and cancer and our measure puts patients into a category of needing a higher-dose scan if they have any concern for cancer in their record in terms of past cancer or suspected new cancer.

So, what makes the measure unique is allowing to do that. What's really important to see in the data is that multiple-phase high-dose studies are used repeatedly when it's simply not necessary, for things like looking for kidney stones, looking for blood clot.

Multiple-phase exams across institutions are used, it depends on the institution but between 7 and 90 percent of the time. So, the variation --

Member Sood: I'm sorry to interrupt but I think in an attempt to keep it focused we were going to just focus on the specific question unless others -- Don is nodding so I think that sounds right.

So, we're just going to focus on that unless Don, you have other questions that --

Co-Chair Yealy: I think the concern was regarding the evidence. Does the evidence show across the entire life spectrum? And I don't think there's a lot of debate about kids. There's exposure versus time kind of issue and I think --

Dr. Smith-Bindman: The Behr report was a report of the National Academy of Medicine, which concluded that these doses are associated with cancers in adults and they quantify the relationship between specific exposures and expected cancer.

So, you can't put together a more seamed panel than they did and that was their conclusion.

Co-Chair Yealy: My goal was just to reframe the question that was being asked. Any other clarity folks

need on the importance of the measure and report, essentially, evidence and performance gap? I see a hand up. Yang?

Member Yu: Yes, I think the NIH and Cancer Institute when they posted information to the general public they do talk about the risks of a high dose of radiation for the general population, not just for the pediatric.

And when we go to a doctor's office, we always get education, not always, sometimes we get education about do you watch out if they offer doses or how often you expose yourself to radiation.

So, I think the risk is there for adult patient populations.

Co-Chair Yealy: Any other either comments or questions?

Co-Chair James: I'd like to ask a question. How standardized are the protocols so that this measure can be used to say you are over the limit?

I can't imagine that there's really that level of standardization from one institute to another and one is an appropriate radiation dose in all the array of specific, clinical diagnostic needs.

Maybe I'm wrong but is there really standardization?

Member Sood: Joel can chime in too, we may be able to opine a tad only from hearing from our clinical radiologists and I don't know if there are other radiologists on the Committee, I don't believe so.

I would love to hear from you if you are.

But my understanding was exactly that, John, there is a lot of clinical variability in terms of why you would choose a study that is higher radiation dose versus not a higher radiation dose that potentially goes beyond the diagnosis code.

At least that was my understanding from talking to

our radiologist. Therefore, it is probably difficult to standardize.

Member Bundy: I think that's what I heard as well, although the ACR has target doses for different study types but there was still variation.

And I think the developers made that clear in their opening paragraphs, that there was I think a 200-fold variation across the country when they looked at different CT scans.

Co-Chair Yealy: I think the developer has shared with us there's a performance gap, I think that's implicit in what's been submitted to us.

Member Bundy: I think from the Staff's standpoint from the evidence, that was rated as moderate and for the performance gap it was rated as high.

Co-Chair Yealy: Other questions or clarifications needed for this part of the measure evaluation?

Ms. Bal: Don, before we jump to voting, quick question, I just wanted to confirm that we're focusing on evidence before we vote and that performance gap will be afterwards.

I heard a couple of mentions of performance gaps so I just wanted to make sure we're keeping it clean.

Co-Chair Yealy: They go right in a row, one after the other. I think we can vote on the next step. And again, we'll be focusing this time on the evidence that was not only submitted but what's been discussed.

Ms. Ingber: Voting is now open for Measure 3633E on Importance to Measure and Report Evidence. Your options are high, moderate, low, or insufficient.

I believe we're looking for a denominator of 17. Just waiting on one more. Because Nancy left I think it might be 16, my apologies.

I will close the poll and share the results. Voting is

now closed on Measure 3633E for Importance to Measure and Report Evidence. We have 1 vote for high, 11 votes for moderate, 3 votes for low, and 2 votes for insufficient.

Therefore, the measure passes on evidence.

Co-Chair Yealy: Geeta, is there anything further regarding the performance gap? I know you've touched upon that already. Anything else that you want to discuss or bring up before we move on?

Member Sood: I would just say the social factors, there really wasn't a differentiator between sexes or races, only from a poverty standpoint.

Those who were significantly in the poverty range had more radiation. And again, as I mentioned before, the performance gap was high as a preliminary rating by the Staff.

Co-Chair Yealy: Any further comments or clarifications needed from the developer?

Member Jackson: I just want to applaud the developer for trying to tackle this problem. In the work that we've done over the past 10 years, I've seen tremendous gaps in the amounts of dosing of radiation across our state.

And I really just think this work is valuable. Is this the perfect way to do it?

I think we can have that debate as we go on through the discussion but this is indeed a big problem and I think there's a great opportunity to improve patient safety in the radiation area.

Co-Chair Yealy: Thanks very much, but I'm still looking for the perfect measure in any domain. So, I'll let you know when I stumble upon that, it's eight years into it.

I don't see any other hands up or chat regarding this part of things so we can move on to the voting.

Ms. Ingber: Voting is now open for Measure 3336E on Importance to Measure and Report Performance Gap. Your options are high, moderate, low, or insufficient and we're looking for a 16 denominator.

Ms. Bal: Hannah, we have 16.

Ms. Ingber: Voting is now closed on Measure 3633E for Importance to Importance to Measure and Report Performance Gap for Measure 3336E.

We have 7 votes for high, 9 votes for moderate, 1 vote for low, and 0 votes for insufficient.

Therefore, the measure passes on performance gap.

Co-Chair Yealy: Thank you, we have another collection of four to discuss the scientific acceptability portion of the measure evaluation.

And again, what I'm hoping the group will do is summarize what's been submitted regarding reliability and validity in that particular order by the developer, summarize any of the Standing Committee survey comments and other comments that exist, and either any other issues that you have either independently or collectively created through your debates.

So, Emily, Teri, Jason, and Sarah, I turn it over to you, I'm not sure who is going to quarterback?

Member Falvey: I know Emily just jumped back on. I'm happy to quarterback while she is getting settled here. I think these conversations will be relatively streamlined.

Member Aaronson: I can dive right back in too, whatever works.

Member Falvey: If you're back and you're ready, I will turn it to you.

Member Aaronson: Sure, I can kick it off. I think we were really happy with ourselves as a team here so I

think everybody will have things to contribute.

And just to confirm, we're right now on the excessive radiation dose or inadequate image quality for diagnostic CT?

Member Hawkins: At the clinician level.

Co-Chair Yealy: And Emily, since I know your day job very well I know you'll be able to very quickly shift gears because that's a skill that's been honed in the emergency department for a long time.

Member Aaronson: That is what I do, exactly.

So, just to quickly discuss the specs, which we'll discuss alongside with the reliability testing, the specs as I'm sure you all discussed a moment ago, this is in ECQM so it looks to assess number of CT scans that are done with high radiation doses.

So, to that end, the numerator is all diagnostic CT studies that have a size-adjusted radiation dose value, which is just the amount of radiation, or global noise value, which is just a measure of image quality, greater than the thresholds specific to the CT category.

And it looks at that in the context of a denominator which is old diagnostic CT studies performed on adults over a one-year period that have an assigned CT category.

And then they also have to have a size-adjusted radiation dose and a global noise value, so they have to have all the components the measure is looking at recorded.

And we'll get into a discussion later I'm sure about missing values which I think is germane to that.

Co-Chair Yealy: Emily, can I make sure we do reliability alone first and then do validity after we voted so they don't get conflated?

Member Aaronson: Yes, fair enough. That was just meant to be a teaser, that's coming later. Those are just the specs that I mentioned.

So, related to reliability very specifically, they perform two types of reliability testing, looked at reliability at the accountable entity level and did signal to noise analysis using interclass correlation coefficient on the EHRs from 16 groups at 7 health systems.

And it's interesting, some of these were midsized groups with 31 physicians, some are quite large, I think as big as 109 physicians. And the mean was 0.99, everyone in their pre-evaluation comments in this group agreed that there were no concerns about reliability and the scientific method panel was also satisfied with the reliability testing for the measure.

And so they rated it as high. So, I think maybe I should pause there.

Co-Chair Yealy: Anyone else from your tag team want to add in? I know that you work collectively.

Member Falvey: I will add one thing that I think will help guide our conversation. I think the developer hinted at this in terms of the denominator statement and numerator statements.

So, the CT scans performed in conjunction with nuclear medicine biopsies procedure related to intervention assessments of bone mineral density are not included in this measure because they're not considered diagnostic CT.

And if I misstated that out of the packet I will ask the developer just to confirm that I got that right at the end here when we're ready for them.

And there were some specific categories and I'm just going to read them so we do it for this one time for all three of these measures.

There's abdomen, pelvis low dose, routine and high

dose, so there's three categories there, cardiac low routine, chest routine and high dose.

Head, low dose, and routine high dose, extremities, neck and cervical spine, thoracic, lumbar spine, and then four considering that are common, CT scans that are done together multi-region.

So, chest, abdomen, pelvis, thoracic and lumbar spine, head, neck routine dose, head, neck high dose.

So, it looks like different standards for bodyweight-adjusted CT dosage that would be expected for each one of those categories.

So, that might help us account for the fact that this might be different for different clinical indications.

It seemed like they thought that through and have kind of a pretty solid collection of common CT scans and some exclusions that we mentioned earlier as potential concerns, that they are not capturing this measure specifically, things that might be in conjunction with cancer-related procedures.

So, I just wanted to point that out to guide our discussion and maybe not go down that road.

Co-Chair Yealy: Thank you. Any other comments, either from the Subcommittee focusing on it or the rest of our Committee?

Member Hawkins: Just to now that there were 606 clinicians at this clinician level and one exclusion just because they had only read one scanner. So, I just wanted to point out that exclusion.

I think that actually raises a good point and a question that I had. It was interesting finding out that they talk about this minimum of 28 CT exams are required to 90 percent reliability.

And I was just wondering how that informed the design of the measure.

So, I'm not sure if we can ask the measure developers anything at this point or if other folks on the Committee have a sense of that or a bit more insight into the significance there?

Co-Chair Yealy: I think that's a good question to pose to the developer.

Dr. Smith-Bindman: Is the question how we came upon the 28 CTs?

Member Aaronson: Yes, and then how you incorporated that into the design of the measure. So, understanding that if I'm interpreting that correctly, you need the minimum of 28 CT scans for this to be 90 percent reliable, then is there a minimum threshold that you're setting for the number of CT scans at the facility level or the individual level?

Dr. Smith-Bindman: I'm going to ask Patrick to possibly answer that question. Patrick, could you weigh into that?

Dr. Romano: Sure, I think it's quite common of course in quality measures that you have to set some sort of minimum threshold for evaluating the performance at the level of the accountable entity whatever that is.

So, we did physical testing that's briefly described in the submission materials to figure out what would be the minimum number of CT scans that we would need in order to meet the reliability thresholds for NQF.

And so that's how we came up with the 28 number. Now, obviously, that number might be subject to evolution over time but the point is the vast majority of radiologists meet that threshold based on their clinical experience.

And so it would really exclude a very small number of radiologists that wouldn't qualify as I think one of the panel members said, just a handful.

Co-Chair Yealy: Other comments or questions, any

need for clarity as we assess reliability right now? Okay, hearing and seeing none I think we can vote on reliability.

Ms. Ingber: Just a reminder as this measure was requested by the Scientific Methods Panel, we're first voting on whether to accept the Scientific Methods Panel's rating for reliability.

Your options are yes and no.

Member Sood: You may have said this and I missed it but what was the Scientific Committee's grading of reliability again? I'm trying to pull it up.

Member Falvey: It was high.

Ms. Ingber: Just one more minute. I'll just give it a couple more seconds because I think someone has rejoined us.

Co-Chair Yealy: What's our target, Hannah, 19 now?

Ms. Ingber: Yes, I believe it's 19 but our forum number is 16, so I just want to give everyone an opportunity. Someone may have stepped away.

All right, I will lock the poll and share the results. For Measure 3633E we have on whether to accept the Scientific Methods Panel's rating for reliability of high, we have 18 votes for yes and 0 votes for no.

Therefore, the measure passes on reliability.

Co-Chair Yealy: Next, we can move on to a focused conversation on validity aspects. Emily, Jason, and Sarah, do you want to review again using the same basic format?

Member Aaronson: Sure, I'm happy to kick it off. So, related to validity, the developers validated this measure at the patient and the encounter level as well as the accountability level.

So, the patient or encounter is synonymous and then

the accountable entity level. So, at the patient or encounter level they looked at a number of different elements that are in the measure, like CT category, patient size, radiation dose, size-adjusted radiation dose and global noise.

And then each of those they looked at and they all performed relatively well. At the accountable entity level they compared the ECQM, so this measure, against medical record review as the gold standard.

And then used a sample of 8000 exams across 8 sites with no discrepancies between the 2.

And it should be noted here there were some public comments about this and I think the most robust was from the American College of Radiology, which raised concerns about the nuance related to the use of NLP-derived data.

And so pointed out that in fact, this was only the comparison done assessing the validity of the NLP-abstracted data itself was actually only done at one site.

And so it raised questions about if that was sufficient to assess. So, really, the face validity, the top felt that it was a relevant metric of quality with 94 to 100 percent of them.

It wasn't clear to me of it was 94 percent or 100 percent of them but 94 to 100 percent of them thought that this would lead to a reduction in CT radiation dose, while maintaining good quality images really importantly.

And there was concern from the Scientific Methods Panel around missing data which is really seen most clearly on Page 105 of the submission, which is Table 2B-3.

And that demonstrates that indeed some sites had as high as 28 percent of missing data from one of the data elements, which was radiation exposure.

But there were no concerns raised in this Committee's comments and the rating for validity was high. So, I'm sure other folks have other things to add to that.

Co-Chair Yealy: Jason or Sarah, anything additional?

Member Falvey: No, I think that was an excellent summary and I think there was very minimal missing data across 0.4 percent.

So, again that concern seemed to be minor and the overall consensus was this had high validity, which was the preliminary ranking from the Scientific Methods Panel as well.

Member Hawkins: Nothing else to add.

Co-Chair Yealy: And Geeta has a question about the risk adjustment falling under the hood of validity. I'm not sure, Geeta, if you want to ask it more directly?

Member Sood: Thank you, sorry to be interrupted but I wanted to ask a little bit more ways to address the variability in clinical factors that probably influence radiation dose.

And one option could I guess be risk adjustment if we understood that other variables other than BMI that would account for differences in what might be considered appropriate radiation doses.

So, I don't know if that's a question or a comment, I will defer to others in the Committee to make that arbitration.

Co-Chair Yealy: I think it's kind of a question so I will turn that over to Rebecca if you want to address that particular issue, and then also you wanted offer clarification about the NLP issue?

Dr. Smith-Bindman: Yes, I think I'll do it in the opposite direction because I think it will be helpful for answering the question.

The measure does not use NLP in any way at all, that was a misunderstanding of our measure within the American College of Radiology Comment. We wrote a detailed response to that. We used NLP at one point within our testing to double-check what we were doing and that's not used by the measure.

The measure entirely uses data stored in ICT10 and CPT codes that are associated with the physician visit when the test is ordered and associated with the radiology exam.

As part of that, we developed an algorithm for using those codes to put patients into what we created, a CT category. There are 19 categories that reflect the reason for the scan that will be associated with dose.

I love the description of our CT categories although there was one subtle error that I'm going to point out because I think it shows what we've done.

So, there's a low-dose CT chest category and a routine dose. For low dose we look at screening for lung cancer or surveillance of lung cancer, routine is the west.

For cardiac we look at calcium scores. So, we've really made very nuanced categories. We published that paper in radiology a few months ago and those categories reflect what we know about the need for exams.

And then we validated the need for those categories and found we were 92 percent accurate. I want to point out that most scans that are done should be done with a routine dose in one phase.

There aren't a gazillion different types of scans that should be done. And the error that many radiologists do is to do multiple phase across a large number of categories, when really, you should use that sparingly.

And so I completely agree that it's important to make

sure patients get assigned to the right category, which we've done and validated, but most patients should be in the routine category as opposed to what we consider exceptions to those categories where you might need a higher or lower dose.

And we've put all imaging of cancer patients into that exception into a high-dose category.

Co-Chair Yealy: Thank you, those clarifications and explanations help a lot. There's one further question but actually, I'm not sure it's a validity question.

Member Pollak: No problem, if it's not validity then no worries. So, I can tell you why I thought it was validity but maybe it's not.

Co-Chair James: Now I'm unmuted. So, if BMI is taken into account, is the age of the patient taken into account? If I'm 85, give me all the radiation you want, I'm not going to be around that long.

If I'm 22, it could matter a lot more to me. Is that considered in the models and in the choice of radiation doses?

Dr. Smith-Bindman: Should I answer?

Co-Chair Yealy: Yes, you may. I wanted to be sure the question was done.

Dr. Smith-Bindman: There is a statistical association of the doses that are used with age but the magnitude of that association is so small that we don't take age into account.

Your question is shouldn't older patients be allowed to have whatever doses there are? And I think there are two ways to answer that question.

The first is that radiologists and technologists use the protocols that are available on the scanners.

And part of the problem in current imaging practice is there are so many choices of protocols that there's

very little consistency in which one gets chosen.

So, if a patient comes from the ED, some technologists pick the first protocol, some the second, some the fifteenth. It's very variable, that's one of the sources of variability.

Having extra protocols available that say, oh, you can use a higher dose in this patient would make it more complicated and there's no evidence that having a higher dose protocol leads to more accurate imaging.

So, since there's no need for it, making it unnecessarily cumbersome I think would push quality in the wrong direction. But the second answer I would provide is actually quite a different answer, which is there is extensive evidence that a CT scan damages the DNA.

A very beautiful paper out of Stanford a few years ago that says one CT scan of a dose over 10 millisievert, which is most scans, damages DNA and activates every marker of DNA damage and repair that are associate with carcinogens.

So, one CT changes that. You might repair some of those changes, you may not repair. And we had no idea what the short-term risk of leukemia is in adults, but the signal from the paper that I showed earlier shows that cancer risk goes up in a few years in adults and if an 85-year-old is otherwise healthy, I believe that would increase that person's risk of cancer within a few years.

I'm not pushing that as the primary driver, I'm pushing primarily let's simplify the process and not have special protocols. But if I were going in for a scan, which I did at Hopkins just a very few years ago, I asked for a low-dose scan.

I'm not quite in my 80s but for my own safety I wanted a low dose, not a high dose, and I think you should the same.

We don't know that you're not increasing your risk of leukemia from undergoing a high-dose CT scan.

Co-Chair Yealy: Thank you. Other questions or concerns?

Member Hawkins: Ed commented on, again, specific body parts being at increased risk. And there was a reviewer that did comment similarly about differentiating between body regions, the head versus the abdomen.

And of course, there may be risk differentiation. So, I think that is where they have a brief discussion if anyone has any insights.

Member Yu: I have a question.

Co-Chair Yealy: Okay, remember we're focused on validity right now, I just want to make sure we don't drift into other areas.

Member Yu: If I drift off, shut me off. I'm not a clinician. Under some circumstances they may require have a higher dose than you would recommend as standard care.

So, I've been reading your measures and do you have a way to really separate this out as a necessity that you have to do it, versus that you're going to measure to capture this is an excessive dose?

Do you have to review the mental records of the individual patient's situation? Is this a validity question? To me that's a validity, whether you can do this.

Co-Chair Yealy: Rebecca, you can certainly answer that.

Dr. Smith-Bindman: I think we demonstrated our capacity to assign patients consistently to these categories. Actually, the person who led this work was Patrick Romano and his team.

Patrick, do you want to talk about the process whereby we determined whether a patient should get a low dose like for a kidney stone or a high dose like for a dissection?

Dr. Romano: Sure, as I think you've described, it was an extensive process that involved both in reviewing the papers in the literature as well as input from experts from different specialties.

For example, we had a number of discussions with cardiologists, urologists, et cetera to really delineate what are the specific things that you would be looking for that would require a higher dose?

And so that was used to create these strata for a low-dose chest scan versus a routine-dose chest scan and so forth? So, I'll also just say that I'm very interested in Geeta's comments about other clinical characteristics of the patients.

I'll say that we looked at this pretty extensively and the only thing we found consistently that was associated with higher doses that wasn't part of the process of care was the patient size.

And that's intuitively clear that for a larger patient, a larger structure, you need a higher dose to get through and image it correctly.

But otherwise, we looked at age, we looked at sex, we looked at a variety of other patient characteristics and just didn't find much that was outside the causal pathway and made a difference.

Member Sood: Thank you. Sorry I'm jumping in and I'm not supposed to but I wanted to just follow up on that. Patrick, thank you for that explanation. That's very helpful and exactly the concern I was trying to get at.

I guess my only follow-up question to that, if it's okay, Don, to ask the developer, is you had mentioned some of the other factors that you had

looked at, the age and sex and other things.

But what about things like tumor size? Because our neurosurgeons will tell us there's a lot of unmeasured variables that are hard to quantify that determine how complex a procedure is or how much detail you need.

Could you just explain, if it's okay, was this chart reviewed? Can we feel good about the fact that there aren't clinical care factors that are influencing the amount of radiation?

Dr. Romano: I might toss it back to Rebecca but I think, obviously, when you get into those kinds of details, it's not feasible to incorporate that into a risk adjustment approach.

But the point is, as Rebecca mentioned earlier, everybody with cancer or suspected cancer gets into the category where they're committed to have a higher dose.

Now, certainly, Rebecca mentioned there is an 8 percent misclassification rate and so we have to be honest about that.

There are some individual cases where there may have been a justification for a higher dose and we tried to err on the side of putting those cases into the higher-dose category, but there might be some misclassification.

But the point is that it's a relatively small number compared with the overall numbers of CT scans that clinicians and facilities are reviewing.

Rebecca might want to clarify a little bit more from the radiology perspective because I'm just a measurement geek.

Dr. Smith-Bindman: No, I think that's correct.

I'm not so sure at UCSF our neurosurgeons need really high doses but our transplant surgeons insist

on incredibly high doses and they want a whole bunch of phases for the way they do their work.

There's no one in radiology who would ever push back. If you need those doses, that's appropriate.

Our categories were created to try to be incredibly generous every time anyone who was in our panel said, oh, we might need higher doses in that setting.

And we tried to put those in the higher-dose category. But the level of individual preference obviously will influence that.

I think current practice is that the inappropriate extreme, where if a doctor says I know this is just a routine abdomen scan but I'd like to routinely scan with four phases through the entire abdomen and pelvis, that's hard to justify.

And it impacts a huge number of scans because abdomen and pelvis is our most common category and routine is our most common category in that.

And that's where most of the problem occurs, where some radiologists just choose to do a four-phase scan without necessarily thinking about it so much. And this measure would suggest is that really necessary?

All of those exams are really above the range and that's sort of what our measure is trying to do, to provide feedback for those common indications which are really driving the radiation overdoses.

Rather than some cases where there's nothing to do but do whatever the clinician needs in that setting.

Co-Chair Yealy: I thought more and have gotten a little Tweet about your question, maybe it is kind of a hybrid of validity and a different area. You want to ask now?

Member Pollak: I think I've kind of gotten the answer so it's not worth going over. The question really was wouldn't a more targeted measure that focused on

those scans which are most commonly overdosed, wouldn't that essentially identify the issue better?

Which is a validity question, the issue at hand.

And I think the answer I'm going to get, because I'm hearing it over and over, is that number one, there's not a perfect measure, and number two, it sounds like in general institution to institution there's going to be variability about which scans.

But if somebody can answer it just succinctly about why they chose to do a global measure including basically all CT scans?

Dr. Smith-Bindman: I think that's a really interesting question, and I think relevant.

And I think your explanation of what I was going to say is correct, which is that it's remarkable how variable institutions are in areas that have radiation doses that are too high.

I think when I became most aware of this issue was a number of years ago I led a randomized trial looking at CT scanning for kidney stones. The study was comparing ultrasound versus CT, which is better, and I'm an ultrasound doctor.

I thought ultrasound would be the right answer, it was 15 hospitals and it was my first randomized trial so I was obsessed with all of the details.

There were a lot of details taken into account. The detail I didn't take into account was the need to tell the radiologists at these 15 institutions that were recruited from the emergency department so I had an ED lead and a radiology lead.

I didn't tell them to use low dose because imaging for kidney stones is one of the areas we know you can use low doses, it's really established and I assumed everyone knew that so I didn't mention it.

It was a huge mistake on my part because when we

looked at the doses we recruited 2500 patients prospectively across 15 centers including Mass General, UCSF, and the doses we used for CT scanning, there were only 5 percent of patients that had the right doses.

95 percent of patients had doses that were really, really high and I put together a meeting where I pulled together the leads at all the hospitals to understand why.

And truly, every 1 of the 15 hospitals had a different reason why their doses were too high. So, if you were correct that we could only focus on those one area that are problems, we would have done that.

But in fact, for some hospitals the dose is a problem for low-dose studies, at others it's a problem for high-dose, at others it's a problem for kids or adults or elderly, or leukemia imaging or appendicitis imaging.

So, we thought the most comprehensive way to look at it would be to look at all CT imaging and then we would capture all of the areas where it's a problem.

And I want to emphasize in that example I gave for CT scanning for kidney stones, there were a couple institutions out of the 15 where it was the radiologists that made a bad decision in my mind.

They used a routine study instead of low dose or high dose.

But for the rest, the radiologists had nothing to do with the decision, some it was the emergency department who said I want to do this kind of imaging or others where they didn't have access to the low-dose protocols.

Others where technologists weren't in the loop and made the wrong decision, so it was really fascinating there was no simple point in the process of decision-making and that's why we want to include the whole process.

Radiologists are different a piece of the puzzle, the system is a piece of the puzzle, the order of providers are a piece of the puzzle.

Co-Chair Yealy: So, Rebecca, you get on, I held my personal question until the end hoping it would get covered. This measure is at the clinician level in contrast to the following measures I'm going to talk about.

And you just hit on my validity concern when it comes to the clinician level because it's which clinician? And are there any concerns about attribution threats?

Because you just named a very common scenario, the bedside physician, the radiologist, and the tech, and the opportunity for pristine choices or less than pristine exist amongst all three.

At the facility level, I know we're not looking at the facility measure now, this is the clinician level, are there any validity issues in attribution?

Dr. Smith-Bindman: I think it's a very good question and certainly, Geeta's question about what a particular provider wants might put tension against what the radiologist wants if the radiologists are being judged but the provider is making those demands.

So, I acknowledge there's definitely some tension there.

I think that greater understanding about how the doses compare for that group of patients to others in the thresholds in our measure will be really helpful to the radiologists, both for their own work to make sure they're consistent in using the right protocols, but also in their discussing the issue with those referred providers to say this is the standard for this, can we try to move that way? And move that way in radiology typically means let's try using a lower-dose exam in a patient and if it doesn't work then we'll shift back.

But I think it empowers the radiologist to then have those difficult discussions.

But the reason we pushed forward with a measure at the facility level was precisely to capture all those process measures so the radiologist has the main concern.

The radiologist is responsible for reading the study and ensuring it's being done in a safe way, that's their main role but definitely has to also reflect the other processes.

In the end, the radiologist gets paid for that exam, they are the ones who say we've done it the right way, this is what we're billing for it.

So, I think they really need to be accountable but understanding there are other people who also influence it who also should be part of that evaluation process.

Co-Chair Yealy: So, as we focus just on the clinician level, what you're saying is the attribution might not directly attach to the most opportune site of change, assuming the opportunity exists.

Dr. Smith-Bindman: No, I think that is the most opportune site of change, that is where the rubber hits the mat.

Co-Chair Yealy: But two of the three people aren't part of this measure, the tech and the person at the bedside isn't the only -- it's attributed to the radiologists, so they indirectly loop back.

That to me is how I would see the measures differently, how a facility one would differ from a clinician-level one. I wouldn't have that same concern, I know we're not talking about the facility one yet.

But under this particular construct that's where I would struggle a little bit. Any other questions under validity for the developing team or a conversation

amongst the group?

Member Aaronson: I just have a clarifying question because I think, Don, the issue you raise is a really important one around attribution, especially as we think about the opportunity here, which is this measure has been presented to us in several different ways.

And so I think right now really focusing on the validity of this particular one at the physician or clinician level.

And to be clear, the attribution, is it the radiologist that protocols the image or the radiologist that reads the image, understanding that at large academic centers like at mine those are different people?

Dr. Smith-Bindman: We, reflecting your question, believe strongly that the best way to measure the performance is at the radiology group level, that would cover all of the radiologists.

But the way the measure is written at the individual level is that the radiologists who bills for that exam is responsible for quality. And the quality of that exam in terms of how it's performed gets attributed to that radiologist who bills for that exam.

I think the disconnect that sometimes happens between radiologists who protocol and radiologists who read it has others problems as well. For example, a radiologist may get an exam in front of them and they don't like the quality of it.

And so they have to work with the doctor who protocols it to ensure that it's done in a way that they are acceptable.

Similarly, if the radiologist is going to be held accountable for the exam, they're going to have ensure that the doctor who protocols it does in such a way that they can take responsibility for quality.

Co-Chair Yealy: Other questions or conversation? I

fully anticipated that the very first measure through would have the most conversation and we would get more and more focused as things went along.

Okay, I think we can vote on validity now at this clinician-level measure. Remember, the scientific panel rated this as acceptable and I have to remind myself I think it was moderate, if I'm not mistaken.

Hang on one second.

Member Falvey: It was high.

Co-Chair Yealy: My error, high.

Ms. Ingber: Voting is now open for Measure 3633E for whether to accept the Scientific Methods Panel's rating for validity as high. Your options are yes and no.

I will share the results. So, for 3633E for whether to accept the Scientific Methods Panel's rating for validity as high we have 14 votes for yes and 3 votes for no, therefore, the measure passes on validity.

Co-Chair Yealy: Thank you very much. Next up is to look sequentially at feasibility, use, and usability. So, if we could focus first on feasibility alone?

Elisa, Brett, Lara, and Ed, you are the quartet that did this. I'm not sure who is going to be quarterbacking this but who would like to lead the evaluation and presentation?

Member Jackson: I think I'm going to kick it off.

Member Charbonneau: I will hand it over to Brett to kick it off. Thanks, Brett.

Member Jackson: And then my colleagues can jump in. The summary is very brief. For feasibility the data is generated through the normal care patients. There are no undue burdens on providers.

The data comes from electronic sources so it's easy

to put together. There are no issues of accuracy and the preliminary rating of feasibility from the Staff is high.

Co-Chair Yealy: Anything else from the group that focused on it?

Member Charbonneau: No.

Co-Chair Yealy: Any questions or clarifications needed from the developer? That type of brevity usually bodes well if I was sitting in a developer's seat.

Member Collins: I had a question. It's my understanding this is commercially available software. What is the feasibility for clinicians or clinical groups that do not have this software or pay for this software?

Did we discuss that?

Member Charbonneau: I think we were going to discuss that in the usability.

Co-Chair Yealy: Yes, I think that is a usability feature, recognizing how these things overlap. Any other questions? Seeing none, I think we can vote on feasibility.

The preliminary Staff rating was high for feasibility.

Ms. Ingber: Voting is now open for Measure 3633E on feasibility. Your options are high, moderate, low, or insufficient. I believe we have all the results.

Voting is now closed on Measure 3633E for feasibility. We have 13 votes for high, 3 votes for moderate, 2 votes for low therefore the measure passes on feasibility.

Co-Chair Yealy: Next we can move on to the use conversation, again this is another must-pass criterion. Initially, the preliminary staff rating was that it passed.

I'll turn things back over to the subgroup that did this. Brett, I don't know if you're also quarterbacking this?

Member Jackson: I think I'm leading this one as well. This is a new measure so it is not in use currently and it is not publicly reported as of today.

It is planned to be submitted to CMS for use with merit-based incentive payment system.

And I think one of my colleagues is going to summarize what the public comment is around use of this metric.

Member Charbonneau: I think, Ed, were you going to do that?

Member Pollak: Sure thing. So, I think at a high level I'll just say there were a number of public comments supporting it for sure but notably, the American Association of Physicists in Medicine and College of Radiology do not support it.

And I know many others have better technical background in this but there were concerns about usability to identify quality of care issues right here.

And it had to do with global measures of noise and things that are way beyond my technical expertise so I'll stop there. We decided to just give that.

I'd be happy to find out if Alicia or Brett want to comment further. There were responses from the measure developer and I know they're on here as well.

Co-Chair Yealy: My question would be were they commenting specifically on use or had they drifted into usability, their separate metrics?

Member Jackson: There was a specific concern, one of the specific concerns is that there's only one vendor, to the question earlier, that is able to pull these reports and capture this data.

Co-Chair Yealy: That sounds more usability to me then. In other words, is it or can the information be pulled and then the next step would be what's the ease around that and the practical issues?

Member Jackson: Yes, there were concerns about being able to do it at scale given that it was done at something like seven health systems I think I saw.

Co-Chair Yealy: Other comments or insights?

Dr. Smith-Bindman: Do you want me to address that?

Co-Chair Yealy: Yes, please. I was waiting to see if there was going to be anymore because they're pretty narrow criterion, I didn't want it to become staccato. But go ahead now.

Dr. Smith-Bindman: In terms of reporting on the measure, currently one vendor has been created. Together with UCSF, we created a company called Alara Imaging that is able to and willing to do that.

But I want to emphasize the measure specifications have all been published in the public domain, physicians and hospitals can report on the measure at no cost using various approaches, be it on their own computers or online using a web interface.

And if other vendors want to develop the capacity to report on this measure, they are welcome to do so. We would welcome more people in this field as possible.

And so the availability of Alara Imaging came in being because of a requirement that CMS put upon us to create an approach that could be used for reporting.

In the process of our measure development, we actually saw guidance from our TEP when CMS told us basically in the middle of the process that we had to step forward in doing this and managing nationwide implementation and reporting, which is beyond the scope of my academic team at UCSF.

And when no group presented itself, we worked to create this opportunity but we're not trying to disallow others from providing solutions to do so.

And just to add one more piece of information, all the data we access for this measure are readily available in the medical records.

What we have done is pull together data that's sometimes stored in the EMR, sometimes is stored in billing claims, sometimes stored in PACs, or the radiology information system.

So, we've brought those data together but these are all very frequently used data elements that we pull together and thus, not outside the possibility of another vendor stepping forward to make use of these data for reporting on our measure.

Member Charbonneau: I guess I have a practical question about that because just from my own experience in a completely different field, even though you're saying that others could come forward and develop this type of software to extract this data, in practical terms, is that realistic to expect of smaller hospitals or smaller hospital systems that may not have a robust IT department to support that kind of development?

How will we make sure we're getting what we need to get from those types of institutions or smaller hospitals would be my concern.

Dr. Smith-Bindman: Super important, thank you for asking that. My simple answer is these smaller institutions can go on a web service and get the data they need without --

I was not suggesting they create these linkages, I meant a new company could do that, but the individual hospitals won't have to.

But I would love to turn this over to either Nate or Simon and Alara to answer your question at a more

technical level than I can do.

Mr. Mazonson: Thanks, Rebecca.

So, I will be repeating some of what Rebecca said but we've created a free and flexible approach to allow sites, whether they are the largest institutions in the country or some of the smaller institutions that you're noting, to be able to report on this measure with low burden.

And again, they are welcome to but not obligated to use the software that has been developed for this express purpose.

Member Charbonneau: But how difficult is the interface between the software and the electronic health record? Is that something that Alara assists with, developing that interface?

Mr. Mazonson: Yes, it is something we assist with to make it as easy as possible for the sites to report on the measure.

Co-Chair Yealy: Other questions, concerns, or insights?

Member Charbonneau: So, just to go back to that again, just because this is something I dealt with in my own system, if a hospital is having technical difficulties does Alara provide tech support for that interface around the clock?

Mr. Mazonson: Alara does provide around the clock technical support to support those sites.

Not only that, if they do work with us to report, we also provide detailed analytics related to their performance on the measure and can help them improve their quality of patient care.

Member Charbonneau: And is there a cost associated with this for the hospital?

Mr. Mazonson: No, we felt it's important to make this

freely available to sites so there's a free option they can use to report on the measure.

Member Charbonneau: Thank you.

Co-Chair Yealy: Other concerns or questions? If not, I think we can vote on use, it's a must-pass criteria and the preliminary rating was passed.

Ms. Ingber: Voting is now open for Measure 3633E on use. Your options are pass and no pass. I'm sorry, I'm having some trouble screen-sharing.

I believe one Committee Member has joined so I'm just making sure we have the right denominator. I'll close the poll and share the results. Voting is now closed for Measure 3633E on use.

We have 17 votes for pass and 1 vote for no pass, therefore, the measure passes on use.

Co-Chair Yealy: And now for our quartet to discuss usability.

Again, what the preliminary Staff rating was, it was high, the notes from any of that meeting as well as any of the feedback from the Committee use and from the public comments and any other concerns that you might have or questions.

Member Jackson: Don, I believe the Staff recommendation was moderate.

Co-Chair Yealy: I thought I had high here, it might be my mistake. Hang on a second, let me cursor out.

Member Charbonneau: It's moderate.

Co-Chair Yealy: I'm sorry, I clicked over to one page off, my mistake. You're right.

Member Jackson: As this is a new measure it is not used currently in any quality improvement program. As we just talked about, the Staff recommendation was moderate.

Predominantly in the comments around usability, there was no concern although I would say that those that expressed concern probably have a higher technical level of understanding of this particular area than maybe the rest of us.

So, that's the summary that I've got right now.

Member Charbonneau: There was some concern that the reduced radiation doses would result in less robust images but I believe the developer found that was not the case.

Member Jackson: Could the developer talk about the use or need for additional scans because of low-quality scans?

Dr. Smith-Bindman: I can. There are two ways that there can be unintended consequences.

One is that the image quality is just not as high as radiologists might want and I guess the second is that could lead to a short-term immediate need to repeat the scan.

In general, the number of exams that are considered of poor quality in radiology is extremely low.

So, in the quality study that we did specifically to understand the thresholds, we created a set of 740 CT scans that we really selected with over-representation of low dose, thinking that's where there would be quality issues.

We had 125 radiologists read those exams for a total number of 25,000 interpretations.

In that set, there are only 3 percent of exams that were graded as having poor image quality, even though we selected exams that would be over-representative.

There was another 8 percent that were considered moderate, not poor but not acceptable. And so poor image quality is a relatively infrequent occurrence in

current imaging but the concern is that if we incentivize lowering the doses, that could lead to a problem in the future.

And so that's the main reason for having the measurement. And we are very aware of the potential for any measure to lead to unintended consequences and this could lead to inadequate images in radiologist.

We don't think that will be a problem, the radiologist's role is to say is the quality acceptable and if not to demand higher quality images. But we know that this could happen and we noted the expression of concern in the letters.

But I want to emphasize we are going to focus on this very heavily to make sure that if there's any detriment in quality, we may have change our thresholds.

The purpose of the measure is really to focus on dose but we obviously don't want to incentivize poor-quality scans.

But in terms of repeating scans, if a scan needs to be repeated for an abdomen scan because it's not good enough and then it happens again that a scan at this location is not good enough, that will be a local indication to the radiologist they have to increase the doses that they're using.

And that's currently what happens in the day-to-day operation of radiology. There's a constant discussion about whether image quality is good or not good and that would continue even if our measure was incentivizing lower doses.

So, we would closely monitor noise and other measures of image quality in the process of assembling data going forward.

In our testing of the data across I think it was 50,000 scans we collected as part of testing, the numbers

that were inadequate of image quality was a small fraction of 1 percent. So, very, very few exams in clinical practice are judged as having poor quality currently.

Co-Chair Yealy: Other questions or concerns?

Member Sood: Not a question but just comments for the developer. That's terrific that you're looking at unintended consequences, thank you for doing that.

I guess I would just broaden the definition of what you're considering, repeat scans, because it may be a couple days later, it may not be an immediate event and it may not be determined by the radiologist.

If the surgeon or whoever doesn't get the information they need, they may request a second scan so that would be my only suggestion.

Co-Chair Yealy: Why don't we move on to voting on this final portion of the measure? And again, the preliminary rating was moderate.

Ms. Ingber: Voting is now open on Measure 3633E for usability. Your options are high, moderate, low, and insufficient. We're just waiting on one more.

Ms. Bal: We're at 18.

Ms. Ingber: Voting is now closed on Measure 3633E on usability. We have 2 votes for high, 15 votes for moderate, 1 votes for low, and 0 votes for insufficient. Therefore the measure passes on usability.

Co-Chair Yealy: Does that move us to our final step, Hannah?

Ms. Ingber: Yes, unless the Committee has any other clarifications, questions, discussion?

Co-Chair Yealy: It's time to make an overall judgment about this.

Ms. Ingber: Voting is now open on Measure 3633E for overall suitability for endorsement. Your options are yes and no. Just give me one moment. I think someone may have rejoined.

Apologies, everyone. Thank you, everyone, for your patience.

Voting is now closed on Measure 3633E for overall suitability for endorsement. We have 15 votes for yes and 4 votes for no, therefore the measure is recommended for endorsement.

Co-Chair Yealy: Thank you all. I propose to the group that we begin --

Member Hawkins: Don, you're on mute.

Co-Chair Yealy: I'm sorry, am I on mute? It doesn't look like it on my end.

Ms. Bal: We can hear you, Don.

Co-Chair Yealy: I propose to the group that we move right on to the next measure.

As we all know, these are incredibly related and I'm hoping now we'll start with the developer for a much more focused presentation, particularly how this differs, pointing out what's different about the measure.

There's one obvious big difference in how it's packaged together and what other key aspects because much of the underpinnings will be exactly the same otherwise.

So, I'll turn it over and if you could just summarize as pithy as you can how this as a standalone measure stands up and how it's distinct from the previous measure.

Imaging/UCSF)

Dr. Smith-Bindman: And just to confirm, this is at the clinician group level?

Co-Chair Yealy: Right, 36, 62E at the clinician group level.

Dr. Smith-Bindman: I think the main advantage of this is that radiologists typically work as a group. As was pointed out by Dr. Aaronson, sometimes one doctor is protocolling and sometimes another is reading.

Often, there are just a handful of radiologists who really take the lead in the area of creating the protocols and it just seems like it makes sense to have the radiologist evaluate it as a group rather than as an individual.

I think from a sample size perspective, individual radiologists read enough studies that you could numerically and statistically summarize their performance but it will be reflected as the performance in the group.

So, we think it really makes sense to summarize the performance of the entire radiology group rather than the individual radiologist. So, I think it's the most obvious way to attribute the performance of CT scanning to associate it with a group of radiologists who typically work together in this area.

Co-Chair Yealy: And aside from the level of which the assessment is made, do you think there's any other important distinctions that the Committee should be aware of in advance?

Dr. Smith-Bindman: I don't.

I think driving quality, both within their group and with all of the outside participants, the hospital, the technologists, the medical physicists makes the most sense as well for that to be driven by the group of radiologists rather than by an individual radiologist.

So, it just seems like it's clearly one that will reflect what they're all doing together.

I think the individual has some strengths but I think the group is, from my point of view, it really makes the most sense in terms of incentivizing and optimizing doses, but providing flexibility in how the group chooses to do that as opposed to laying the blame or benefit on a single individual radiologist.

Co-Chair Yealy: I appreciate that. I guess I'd ask our quartet now assigned, that's Joel, Nancy, Geeta, and Yanling to walk through initially the evidence behind this. Again, I'd like to keep this focused.

We spent a lot of time on the previous measure, we don't want to short-shrift this but we don't have to revisit issues that have been extensively evaluated before.

Member Bundy: This is Joel. I think the evidence presented was the same as what we discussed in the last measure. There really wasn't anything in addition, and the concerns we had were around pediatrics.

I think we walked through those last time.

Co-Chair Yealy: Others from the group, at least the quartet, and then we can go back to the whole group?

Member Yu: I don't have anything new to add, Joel said it well.

Member Schoenborn: I don't either.

Co-Chair Yealy: From the rest of Committee, any questions or concerns? Hearing none, I think we can vote on this part of the evaluation.

Ms. Ingber: Voting is now open for Measure 3662E on Importance to Measure and Report Evidence. Your options are high, moderate, low, or insufficient.

We're just waiting on a few more.

Co-Chair Yealy: Hannah, for the group I'll point out that if I click your slide, nothing happens, for those of you who don't have dual screens up and forget to refresh to the other screen.

Ms. Ingber: I think we're waiting on just one more. I'll just give it a few more seconds. I will lock the poll and share the results. Voting is now closed on Measure 3662E on Evidence.

We have 0 votes for high, 16 votes for moderate, 0 votes for low, and 2 votes for insufficient, therefore, the measure passes on evidence.

Co-Chair Yealy: Next is the performance gap and I have the same quartet. And again, the discussion around this I assume will look very much like the previous evaluation.

Member Bundy: The same thing as before, the performance score was 30 percent and the standard deviation of 7 percent across about 43,000 CT scans and 16 clinician groups. And the Staff preliminary rating was high.

Co-Chair Yealy: Any other comments or observations from first the quartet? If not the quartet, the rest of the Committee? I think we can vote on the performance gap.

Ms. Ingber: Voting is now open for Measure 3662E for performance gap. Your options are high, moderate, low, or insufficient. We're just waiting on a few more.

I'll lock the poll. Voting is now closed for Measure 3662E for performance gap. We have 8 votes for high, 10 votes for moderate, 0 votes for low, and 0 votes for insufficient.

Therefore, the measure passes on performance gap.

Co-Chair Yealy: Thank you all very much. We can move now onto the scientific acceptability and we have Emily, Jason, and Sarah to discuss that first and

then we'll focus on validity.

But let's focus on reliability first. Again, I expect the pattern to be similar to, although it may not be identical to the previous conversations. Emily, are you leading this one again?

Member Falvey: I think I will this time, Emily, kindly turn over the couple of sentences that Joe left here for me. Hopefully we can keep this streamlined. The numerator and denominator for the measure is identical to the last one.

Just, for the record, it's all the diagnostic CT scans for the size suggestion radiation dose for global noise value are being collected as a percentage of all diagnostic CT scans that have the assigned category, have the measure developer outlined in their packet.

And the exclusions are the same as well, missing data and those are missing age or dosages from the CT scans.

In terms of reliability, the developers did split sample testing and signal to noise analysis from 16 groups within 7 healthcare systems in a vertically-integrated organization, as done over February 2020 to April 2021.

Those clinician groups ranged in size from 31 to 109 with an average group size of 27, so these are reasonably sized groups. And then it had the split sample, interclass correlation of 0.99 for the data collection period.

And similar to before, the minimum was 28 CT scans, required to achieve the reliability rating. I think we talked about that last time and the overall scientific methods panel for reliability was high.

Co-Chair Yealy: Thanks, Jason, anything else from your partners in crime?

Member Hawkins: No, he's captured it.

Co-Chair Yealy: Any questions or insights from the rest of the Committee?

Member Thraen: This is Iona, I have a quick question.

The 28 case standard, is that applied at the -- in this instance this is a group evaluation, was that applied at the individual provider level or at the group level?

Dr. Smith-Bindman: At the group level.

Member Thraen: So, it was 28 cases regardless of the size of the group?

Dr. Smith-Bindman: That's correct.

Co-Chair Yealy: So, you would assume that it's even less likely a group would fall out. That's a group that's going to have a hard time putting shoes on the baby if 28 images aren't...

Dr. Smith-Bindman: Even tiny practices, if there's a CT machine there they have to at least do a scan a day, it makes sense to have CT even in an extremely rural location.

Co-Chair Yealy: A scan a week would get them in. Any other questions or clarifications? Seeing none I think we can vote on reliability

Ms. Ingber: Voting is now open for Measure 3662E for whether to acceptable the Scientific Methods Panel rating as high for reliability. Your options are yes and no.

Waiting on one more. Voting is now closed for Measure 3662E for reliability. We have 18 votes for yes and zero votes for no, therefore, the measure passes on reliability.

Co-Chair Yealy: Jason and Sarah, do you want to discuss the validity section of acceptability now?

Member Falvey: Sure, I'll go ahead and kick us off and I'll let Sarah or Emily chime in.

The validity piece was identical to the last measure, the validity testing was done at the patient encounter level and then we talked about the accuracy of the CT category based on ICD-10 extensively last time.

And those were identical for this, patient size, radiation dose, size adjustment to radiation dose, global noise were all measured identically to the last measure and the validity testing at the accountable entity organization, they compared against medical record review from the eight health systems and found there was almost no discrepancies between the outputs and the manually collected information.

The face validity information was identical to the last measure as well. There wasn't any serious validity concerns that were raised in the general comments.

And the Scientific Methods Panel's review and preliminary rating was high.

Co-Chair Yealy: Other comments, insights, or any need for clarity?

The last time around we talked about attribution issues in validity and as we move further away from the clinician into the group and then into the facility, those attribution issues change.

And you might even say they dissipate to some degree because at a facility level you're shuffling deck-chairs about people who are responsible.

And so I still have some attribution validity concerns that overlap at the clinician level, but they differ a little bit.

Member Sood: Would you mind sharing, Don, a little bit more about how you're envisioning that? That was a really great point at the clinician level.

Would you mind also expanding on what you were thinking at the group level?

Co-Chair Yealy: The group, depending on the

structure at an individual site, may not include, if we just use there are a minimum of three people involved with this, the bedside provider, the imaging technician, and then the radiologist, the modifier clinician level to group level may or may not dissipate some of those concerns and then at the facility level may or may not again.

It's just as simple as that, and how that impacts the utility of the measure is exactly what everybody's voting on. I'm not here to say that it's right or wrong, it's just that it's a natural question.

At my institution the group level would be the same, it would be the University of Pittsburgh Physicians. So, it would wash away, you would just have the technical versus all the essentially physician providers.

I don't want to say versus, it's an awkward way to frame the issue. Any other questions or needs for clarity from the group? If not, I think we can vote on validity.

Ms. Ingber: Voting is now open for Measure 3662E on validity, whether to accept the Scientific Methods Panel's rating of 5 for validity. Your options are yes and no.

I think we're just waiting on one more. I will close the poll and share the results.

Voting is now closed on Measure 3662E on validity. We have 16 votes for yes and 3 votes for no, therefore, the measure passes on validity.

Co-Chair Yealy: I think that is a testimony to our consistency, isn't it? And now we can move on to the feasibility use and usability evaluations, starting first with feasibility focus alone.

Elisa, Brett, Alara, and Ed, this was in your bailiwick. I'm not sure who's going to lead the discussion?

Member Charbonneau: I think, essentially, we have

a ditto report from the last time so, Brett, did you want to review that again? It's a new study.

Member Jackson: I agree, I'll just echo the same comments as before and just reminding the Committee that the preliminary rating for feasibility was high.

Co-Chair Yealy: Elisa, I never would have thought that you would have channeled your inner Rush Limbaugh by saying ditto but there's a little bit of Rush in everybody, I guess.

Member Charbonneau: No.

Co-Chair Yealy: Any questions, clarity, need for conversation? We can vote on feasibility.

Ms. Ingber: Voting is now open for Measure 3662E on feasibility. Your options are high, moderate, low, or insufficient.

Thank you, everyone. Voting is now closed for Measure 3662E on feasibility. We have 11 votes for high, 7 votes for moderate, 1 vote for low and 0 votes for insufficient.

Therefore, the measure passes on feasibility.

Co-Chair Yealy: Next up we can discuss use, again our same group and I expect an overlapping approach.

Member Jackson: The measure is not currently in use or used in an accountability program and the preliminary rating for use is pass.

Member Charbonneau: -- does have plans to submit to CMS for MIPS and MUC, which is the incentive plan and measures under consideration for 2022.

Co-Chair Yealy: Other questions, concerns, need for clarity? Hearing none, I think we're ready to vote on use.

Ms. Ingber: Voting is now open for Measure 3662E on use. Your options are pass and no pass.

We're just waiting on one more. Voting is now closed for Measure 3662E on use. We have 18 votes for pass and 1 vote for no pass.

Therefore, the measure passes on use.

Co-Chair Yealy: Next up is usability, again, any comments from our subgroup that's targeted to evaluate this portion of the measure?

Member Pollak: The public comments were copy and paste so it's the same thing.

Co-Chair Yealy: You mean like every resident note I read or something different than that?

Member Charbonneau: Ouch.

Co-Chair Yealy: Gross generalizations all exist because they're true, right? Any need for clarity or conversation? Hearing none, I think we can vote on usability.

Ms. Ingber: Voting is now open for Measure 3662E on usability. Your options are high, moderate, low, and insufficient.

We're just waiting on one more. I will close the poll and share the results. Voting is now closed for Measure 3662E on usability. We have 2 votes for high, 15 votes for moderate, 1 vote for low, and 0 votes for insufficient.

Therefore, the measure passes on usability.

Co-Chair Yealy: Now it's time for the final evaluation.

Member Aaronson: Just a question, when do we talk about competing measures?

Co-Chair Yealy: At the end of everything because they have to exist first to then talk about competition.

Ms. Ingber: I'll open the poll on overall suitability. The voting is now open for Measure 3662E on overall suitability for endorsement. Your options are yes and no.

We're just waiting on one more. Voting is now closed for Measure 3662E for overall suitability for endorsement. We have 15 votes for yes and 3 votes for no, therefore, the measure is recommended for endorsement.

Co-Chair Yealy: Thank you all. We have only one more measure to do, the third in this series.

What I'd like to ask the group is, again, I think it's going to be a much more focused and likely shorter set of evaluations given the previous discussions.

We could take a break right now or we could try to get through the measure. I suspect that it would take us 20 minutes or less. I'm open to either one, I want to be respectful of people's need for even just a short break right now.

So, what I'm going to ask, and I think I can see almost everybody, put a thumbs-up and hold it up for me if you want to take a very short break.

I don't have, Hannah, the same tool that you have but I think consensus was reached on this part of the measure. So, why don't we take a ten-minute break?

It is 3:48 p.m. so we will be crisply starting before 4:00 p.m., okay?

(Whereupon, the above-entitled matter went off the record at 3:48 p.m. and resumed at 3:59 p.m.)

3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed tomography (CT) in Adults (Facility Level) (Alara Imaging/UCSF)

Co-Chair Yealy: Thank you very much, and we've certainly had an acceleration in our pace, no doubt about that.

So, we have one final measure to review now, obviously part of our triad of measures to review here. And that's 3663, the same basic CT measure but this is now at the facility level.

I'd ask the developer once again to highlight, aside from the measurement prism now being the facility instead of the clinician or the group level, are there any other substantive differences that you think are important for the Committee to know now?

Dr. Smith-Bindman: The only comment I wanted to raise is to just tell you guys where this additional level of measurement came from because we were originally contracted to develop the physician measure.

And it was really our TEP that helped us move in this direction, raising concerns that radiologists would not always have access to the data if they were basically controlled by the hospitals and that there would not always be clear alignment in how to move towards dose optimization if only the radiologist and not the other players, the most important one being the hospital, were not similarly aligned.

So, they really encouraged us to think about adding the hospital measure to enhance the overall agenda of trying to improve quality, basically to align the incentives and to provide access to the data that both sides would need for reporting purposes.

Co-Chair Yealy: Thank you very much. Without knowing it, you answered the question I was going to ask you at the end of the metric so I won't even need to ask it now.

Thanks, I appreciate it. Let me turn things over now to our discussants.

The quartet for discussing the importance to measure and report that is evidence first we'll discuss, and then performance gap is again Joel, Nancy, Geeta, and Yanling.

Anything you want to add? Since this is now the third conversation with a very similar measure.

Member Bundy: No, Don, I don't think so, it's the same evidence that was presented.

Member Yu: I don't have anything more to add, thank you for asking.

Co-Chair Yealy: And I can see my humor from the previous session eliminated use of the word ditto completely, didn't it? So, I accomplished at least one thing today.

Any questions, need for clarity, anything from the rest of the group? Hearing none, I think we can vote on the evidence part of this particular measure.

Thanks, voting is now open for Measure 3663E on Importance to Measure and Report of Evidence. Your options are high, moderate, low, and insufficient.

And again, we need 16 votes for a quorum but I'll give it a few more seconds just to give everyone who's coming back a chance to vote. All right, let's calculate the results.

And I'll read the results off. Voting is now closed for Measure 3663E for evidence. We have 1 vote for high, 14 votes for moderate, 1 vote for low, and 1 vote for insufficient.

Therefore, the measure passes on evidence.

Co-Chair Yealy: Next is to evaluate and discuss the performance gap. Joel, Nancy, Geeta, and Yanling, anything to share about either the presentation of it, the feedback that exists already, or your insights?

Member Bundy: It was the same, Don, and the Staff preliminary rating was also high.

Co-Chair Yealy: Questions from the group or a need from the developer?

Seeing none we can vote on performance gap.

Ms. Ingber: Voting is now open for Measure 3663E on performance gap.

Member Charbonneau: My screen says the poll is locked.

Ms. Ingber: Thank you, it should be good now. I'm just waiting for one more. Voting is now closed for Measure 3663E for performance gap.

We have 7 votes for high, 10 votes for moderate, 0 vote for low, and 0 votes for insufficient. Therefore, the measure passes on performance gap.

Co-Chair Yealy: Thank you very much. We'll move next into the scientific acceptability phase of evaluation.

Jason, Sarah, and Emily can talk us through initially the reliability assessments, both what's been presented, what the comments have suggested and any other insights from the group.

Member Falvey: I'm happy to lead us off. I know Emily looks like she is off camera but here. I think this is a very similar conversation to before, same numerator and denominator statements and exclusions from this measure.

I think for reliability the differences were simply just testing out the hospital level and they obtained CT scans during inpatient hospitalizations.

So, they range from 134 to 1568 scans at the hospital level and they did a split sample of interclass correlations and very similarly to the other measures we looked at, the reliability interclass correlation coefficient was greater than 0.99 within each hospital.

And the predicted reliability exceeded 0.99 for each hospital during the testing phase. I will point out these are all during inpatient hospitalizations and it

may be worth clarifying that this measure is also being used for scans that are done at the outpatient level.

I cannot imagine they differ that much but I think that would be an important clarification because I believe it's only shown for inpatient hospitalizations.

And if there are technical differences there in terms of scan indications, maybe I can just let the developer quickly comment on that and move on.

Co-Chair Yealy: Other comments or questions before we bring the developer back in?

Member Aaronson: I have nothing to add.

Co-Chair Yealy: Rebecca, do you want to answer the question about the in versus outpatient?

Dr. Smith-Bindman: We would just ask to specify the results in that way, the results are the same.

The indications are not identical so there will be more trauma scans in one setting than the other, more strokes in another, but the results turned out to be identical.

We would just ask to separate it that way.

Co-Chair Yealy: Any other clarity needed from the group? If not, let's vote on reliability.

Ms. Ingber: Voting is now open for Measure 3663E on whether to acceptable the Scientific Methods Panel's rating for reliability.

Your options are yes and no.

Member Falvey: It was high, just for everybody.

Ms. Ingber: I'm just waiting on three more. I think one more, I'll just give it a couple seconds. I'll close the poll and share the results. Thank you, everyone.

Voting is now closed for Measure 3663 E on reliability.

We have 16 votes for yes and 0 for no, therefore the measure passes on reliability.

Co-Chair Yealy: Next up will be the discussion around validity and again, Jason, Sarah, and Emily, anything you'd like to add?

Member Falvey: This is going to be very identical to the last two discussions that we had. The validity testing was done at the same patient encounter level categories.

CT categories, patient size, radiation dose, size of adjusted radiation dose and global noise were all measured and there was no validity concerns around the actual measures or the missing data at the accountable entity level.

There was very good validity between the gold standard of a human reviewed indicator and what the machines were outputting, and similarly the face validity was about to be very high.

And the Scientific Methods Panel's preliminary rating for validity was also high.

Co-Chair Yealy: Other questions, concerns, or need for clarity from the group? I have less of an attribution concern here as we get further up the observation prism as it were.

Okay, hearing none, we can go ahead and vote on validity.

Ms. Ingber: Voting is now open for Measure 3663E on validity on whether to accept the scientific method panel's rating for validity. Your options are yes and no.

I'll just give it a few more seconds.

Voting is now closed for Measure 3663 E on validity. We have 15 votes for yes and 1 for no, therefore the measure passes on validity.

Co-Chair Yealy: Next up is our triad conversation on feasibility use and usability, starting first with a focus first on feasibility. Brett, Elisa, Alara, and Ed, anything you'd like to add or anything to compare and contrast with previous conversations?

Member Jackson: Nothing new to add.

Member Charbonneau: I don't think we have anything different to add.

Co-Chair Yealy: And the previous evaluation of this on feasibility was?

Member Jackson: High.

Member Pollak: Feasibility is high.

Co-Chair Yealy: Anybody need any conversation, clarification? If not, let's go ahead and vote on feasibility.

Ms. Ingber: Voting is now open for Measure 3663E on feasibility. Your options are high, moderate, low, and insufficient. I'll give it just a couple more seconds.

Voting is now closed for Measure 3663E on feasibility. We have 12 votes for high, 4 votes for moderate, 1 vote for low, and 0 votes for insufficient.

Therefore, the measure passes on feasibility.

Co-Chair Yealy: Next up is the conversation on use. Again, our quartet, any thoughts that differ from previous conversations?

Member Bundy: No, sir.

Co-Chair Yealy: Questions, comments, or clarity needed for anyone else on the call or Committee? Seeing none, I think we can vote on use.

Ms. Ingber: Voting is now open for Measure 3663E on use. Your options are pass or no pass. I'll give it just a couple more seconds.

Voting is now closed for Measure 3663E on use. We have 16 votes for pass and 1 vote for no pass, therefore, the measure passes on use.

Co-Chair Yealy: And now the conversation from usability. Once again, from our quartet any new insights from what the previous evaluation was?

Member Charbonneau: No, we're actually a trio but I don't think we have anything new to add.

Member Jackson: And the rating was moderate once again.

Co-Chair Yealy: Questions or needs from the rest of the voting panel? I think we can vote on usability.

Ms. Ingber: Voting is now open for Measure 3663E on usability. Your options are high, moderate, low and insufficient. I'll wait for just a couple more seconds.

Voting is now closed for Measure 3663E on usability. We have 2 votes for high, 14 votes for moderate, 1 vote for low, and 0 votes for insufficient. Therefore, the measure passes on usability.

Co-Chair Yealy: And we can move on and vote on the final overall acceptability.

Ms. Ingber: I think so. Voting is now open for Measure 3663E on overall suitability for endorsement. Your options are yes and no. Just a couple more seconds.

Voting is now closed for Measure 3663E on overall suitability for endorsement. We have 15 votes for yes and 2 votes for no, therefore, the measure is recommended for endorsement. Therefore, the measure passes on usability.

Co-Chair Yealy: Thank you, we're through the bulk of the now. We have to now discuss related and competing measures.

Ms. Buchanan: That's right. Thank you, Don, for leading a very smooth process this afternoon with all the ECQMs and thank you, John, for leading this morning.

I'm really glad that everything went smoothly and everyone adapted well to the different process this afternoon.

So, I think everyone is excited to talk about relating and completing this cycle. Next slide, please.

To provide a brief overview of what's considered competing and what's considered related, a competing measure is the same concept and the same target population.

In competing institutions the Standing Committee would ultimately have a best in class discussion. There are also related measures where they have a different target population and/or a different concept.

If they're both different, we don't have any competition between the measures and no harmonization is needed. If there are some similarities, developers are asked to harmonize their measures with other related measures appropriately.

Next slide, please. Early measures recommended for endorsement will be discussed. The Committee will not be asked to select the best in class measure if all related and competing measures are not currently under review.

The Committee discuss harmonization and make recommendations.

If there are similarities between measures, the point of this conversation is to see if the Standing Committee has any questions or concerns with what the developers listed in their measure submission with regards to related measures.

Or if there are any recommendations they'd like to offer the developers to be noted in the final technical report.

The overall goal is to mitigate any potential burden to the system in the number of measures and differences across related measures.

Next slide, please.

The questions we want you to keep in mind during this discussion are the measure specifications for the related measure harmonized to the extent possible.

Are there differences that could impact interpretability and add data collection burden? And are the differences justified?

So, next slide, please. First, I want to note there are no competing measures this cycle so we're going to walk through all of the related measures identified by the developers during measure submission.

Remember that recommendations will not change the endorsement vote in any way and they'll be noted in the final report for future evaluations by the Standing Committee.

So, the first related conversation will be for Measure 3636. I think we've already discussed the influenza measure already just in talking about Measure 3636.

But the developer did provide some additional background information in the submission. I will read that off for you all. The proposed measure is harmonized to use the same denominator categories as 0431.

The target population of both 0431 and the proposed measure is healthcare personnel who may be encountered by other healthcare personnel and patients during the reporting period.

However the data collection and reporting period is annually for the six months from October to March

for 0431, the data collection period for the proposed measure is one week a month and the reporting period is quarterly every three months.

The developer provided the rationale that the shorter data collection period for the proposed measure is towards the reporting burden. The rationale for more frequent reporting period that is not seasonal is COVID-19 vaccination recovery is a public health priority and COVID-19 has not yet demonstrated consistencies like influenza.

Because of the different time periods for data collection and reporting, NQF 0431 includes healthcare personnel who worked for at least one day during the six months data collection period, while the proposed measure includes healthcare personnel who are scheduled to work regularly.

Many healthcare who regularly work in a facility may be temporarily absent from a facility for periods of up to two weeks due to illness, injury, or vacation/leave.

Because the measurement period covered by the influenza vaccination measure, such absences will not impact the influenza measure denominator.

However, the COVID-19 vaccination measurement period is only a week for each month of the quarter, so a number of regularly working healthcare personnel may be absent during this shortened period.

Therefore, healthcare personnel who regularly work in the facility may be temporarily absent from the facility for up to two weeks are still to be included.

For many facilities, collecting data for workers who regularly work in the facility reduces the data collection burden as a daily accounting of healthcare personnel work hours is not required.

Also, reporting and calculating covering rates for the three-month time periods rather than annually for a

six-month time period is a higher burden but one that is warranted for disease which has and continues to be the cause of a worldwide pandemic.

So, that was the developer's rationale around harmonization. And now I'll turn it over to the Co-Chairs to facilitate any discussion around these two measures.

Co-Chair Yealy: Thanks very much. They're obviously related but distinctly different, and I think the comments about seasonality are just uncertain.

We might be making a different decision in a year or two or three when we know a little bit more about the patterns of COVID-19 as it becomes more endemic rather than pandemic.

I see footprints which would allow much more harmonization and efficiency eventually coming but they're just the footprints right now, is my view on things.

John, I don't know if you have any other thoughts?

Co-Chair James: Don, that's consistent. The jury is not out on how to do measures for COVID-19 yet in the long term. So, it needs to be kept separate.

Co-Chair Yealy: Thoughts from the voting Committee, anything else you think we should be thinking about?

Member Sood: This is Geeta, I agree with what you're saying. We're not sure what the pattern is going to be yet for SARS-CoV and I've discovered the virus is smarter than I am for sure.

And, Don, I think you might have pissed part of what the CDC had said but it sounded like they were going to be flexible with the reporting time based on what would be happening over the next few months.

So, that's in line with what you were saying that there would be some flexibility with that.

Co-Chair Yealy: Other thoughts or input?

Ms. Bal: There's a comment by Anne in the chat.

Co-Chair Yealy: It goes along with we don't know how much is seasonal versus behavioral.

That same comment could be true about influenza though also, we just really have a known experience of a couple hundred years with influenza.

We certainly have it with various forms of Coronavirus, we just have not really sought it out all that often.

Member Thraen: This is Iona.

The other thing I would add and just highlight, the conversation about contract workers and the carve-out in the measure related to contract workers in the change in the employment environment related to travel nurses, aid contracts, provider contracts, et cetera.

COVID has driven a larger portion of those people working in the environments and the measure currently does not include them in the measure.

Co-Chair Yealy: And that's a distinct difference from the CMS mandate, which does not care who your employer is, it only identifies where you're at. So, a vendor who is inside the hospital would have the same requirement as someone who is employed or even an independent contractor, that is an important distinction.

One of the other differences, and I don't know that we can adjudicate it here, is healthcare personnel versus healthcare worker. For not the previous measure but the overall CMS mandate you have to be patient-facing or directly interacting with patient-facing.

So, people who are employed but have zero contact either with patients or with someone who comes in

contact with patients would not fall under that. I don't want to conflate the federal mandate with the measures because they're not exactly the same.

But the reality here is the people who will have to deal with it will be dealing with all three of these at once, federal mandates, influenza measures, and then COVID-19 measures.

And as much alignment will really make things easier on end users and send a clearer message. I'm not sure what the next steps we need with this related measure concern here.

Ms. Buchanan: I'm hearing that there aren't really any recommendations for the developer at this point.

Co-Chair Yealy: I think we move forward and move to comment at some time in the future.

Ms. Buchanan: Yes. So, next we have queued up to discuss the ECQMs against the measures that are listed in the measure submissions.

But we've wanted to hold this space first to allow you all to discuss the three ECQMs against each other if you'd like to. Since it isn't part of the submission, we don't have developer rationale laid out from the submission. But I wanted to open the floor, we've already talked about the differences between the three measures but any recommendations and any thoughts about the three points we want to consider with related or competing measures?

Co-Chair Yealy: So, in order to prompt the conversation along and based on the conversations that we've had, I guess I would have a question for the developer and it's going to be a construct.

And I just ask you accept my question on face value and answer it and say if you had to pick one of these, which measure would you pick and why?

I know what change you're trying to do with patient-focused downstream change but if you had to pick

one of these between the three measures, putting aside for a second the pediatric component, which one would you pick? And then I'm happy to share what my views on it are. You can probably guess.

Dr. Smith-Bindman: I'm guessing what your view is and I understand the question. I use a question like that all the time in my personal life and with my family, I use it at work.

So, I think it's a really good question that I just don't feel able to answer. I think that having the different measures enhance one another immeasurably and having the individual and group measures really provides incredible flexibility that users of the measures, be it payers or regulators, could use as appropriate.

So, I think having them all really allows them to enhance each other.

I think that part of the challenge in this space is the lack of complete accountability, so it's easy to point fingers at the hospital for not buying the fancy software you want, or the radiologist for not managing the protocols like you want them to.

And I think having everyone work together to be similarly incentivized will have a larger effect that will magnify the effect of any of the individual ones.

So, I think clearly in this space the radiologist is the Captain of the team, the quarterback, but I think radiology has become and is really a team sport.

Everyone is in it together. And so I think all of these measures really have their role. I think they're all really important and would enhance each other for all of them to move forward.

Co-Chair Yealy: Thoughts from the group?

Member Sood: I'm going to just echo what, Don, you and Emily had said, who might have popped off already, but I think you're very right that the validity

concerns at a clinician level are concerning.

They're not the ones that are truly accountable in the same way as the group or the facility. So, I think I will amplify what you had said and I think I will vote for facility level for these.

That seem to be the most fair accountability-wise.

Member Pollak: I agree totally. I think all the time we hold facilities accountable when obviously, they then have to work with their individual team members and in many cases it's much more herding cats than this in my opinion.

Member Thraen: This is Iona. Just to add to that, I think the radiology being the captain of the ship based on what the earlier conversation is, and what we've seen, is true to an extent, but a lot of time, it's the oncologist driving the request, it's the trauma surgeon driving the request, et cetera.

And so the radiologists are often responding to that request or desire or pressure, so it's really at the facility level that you have the opportunity I think to bring all the players to the table, whereas just radiology on its own trying to set standards without engagement with these other key decision-makers I don't think is as fruitful.

Co-Chair Yealy: Other thoughts?

You know, Rebecca, one of the opportunities again would be -- I'm not sure there's an easy path to this -- would be to find a way to combine this.

What we're hearing is there is redundancy built in, and you saw that as a value and the question is, what's the incremental value for is the juice for every overlap worth the squeeze?

And so I framed at the extreme, where do you get the most juice if you had to look at them individually?

That's false framing, I get that. That's meant to be

an intellectual exercise.

Dr. Smith-Bindman: No, I understood that. Can I add one point?

Which is, there's both an issue around drivers of quality, which is what some of the comments are about, which is the radiologists can't drive quality alone.

And I made comments in the chat that half of CT scans are done outside of the hospital setting or facility entirely, so those need to be covered.

But the reason we developed the parallel hospital measure was because of practicality, and on a practical level, the data that are needed to calculate this eCQM are not uniformly under the control of the hospital or the radiologists.

So the EHR might be under the control of the hospital.

The images stored in PACS may be the collection of the radiologists, and if both are not required or incentivized or motivated to share the data, it may be impossible for hospitals to get these data without the radiologists being on board, and it may be impossible for radiologists to get this data without the hospitals being on board.

The marginal additional work involve is non-existent. There is no additional work involved in assembling the data. So it permits attribution at different levels with the same amount of work.

So I think there is a very strong need for the measures at both the clinician and the facility level, but it's not just a matter of enhancing or amplifying, but also on a practical level it's just not feasible if clearly one of those groups is written out of being evaluated.

Co-Chair Yealy: Jason just commented on something that triggered in my head when you were speaking.

Is it possible to combine the clinicians and the groups together, and then have a combined measure that looked at facility and groups to achieve all of those goods that you're saying? Although you're --

(Simultaneous speaking.)

Co-Chair Yealy: -- exact same amount, three separate measures, no matter how you slice it is three different --

Dr. Smith-Bindman: So I want to just answer that question about individual over a group.

CMS has been very explicit that they want the measure at those two levels, and they absolutely insist they have built into their new submission to the MUC list for this round, published a few weeks ago, that you need to report the data at the individual clinician level.

They won't accept it otherwise. And --

Co-Chair Yealy: Right. I think we get back though to the original, though. At the clinician level, you have -- or found attribution problems.

Dr. Smith-Bindman: I respectfully disagree.

Radiologists made so many stands, it's not a perfect reflection of everything that could be done, but I think the individual radiologist who does take attribution when billing for that exam needs to also take attribution of the quality or inequality gaps that comes along with doing something that you bill for.

You have to ensure it's being done the right way and the safe way.

Member Charbonneau: So can I ask a question as a non-radiologist that I think might help some of us maybe wrap our head around this?

So, let's say an x-ray is ordered in the ER and the radiologist is reading the x-ray to rule out a fracture,

and says, I need another view because I can't see well enough to definitively diagnose this fracture.

Would that be analogous to the responsibility of the reading radiologists of a CT scan saying the quality is not good enough for me to answer this clinical question, so that's why you're feeling that that's where the attribution should be?

That it's up to that final radiologist who's reading the scan or the x-ray to make sure the quality is sufficient?

Dr. Smith-Bindman: If I'm understanding what you're asking, I think that's a good analogy.

The quality of the image itself, not the appropriateness of the image -- should I have gotten an ankle image because the person actually hurt their ankle, not their hand, right?

That's not what the radiologist is responsible in this measure. They're responsible for I've done this kind of scan, was it done in the right way?

And I think I could generate examples when that might be challenging.

So one such example might be the radiologist in their system doesn't know why the study was done, wasn't provided with that information, and therefore they can't tailor the way the radiation dose should be tailored to the clinical indication. That's a quality problem.

The radiologist is legally responsible for knowing why the study was done, what the clinical question is, and so if in their system that gap exists where they don't have that information, that that needs to be fixed, and this measure could drive an improvement in that, and say we got to figure out how to do this, and this is a longstanding problem in radiology that many groups have figured out different solutions, but the person who's making decisions about how to scan the

patient needs to know why the patient was scanned, but they're responsible to make sure that what was done was done in a way that lead to useful information.

So I think that's a good analogy, and the person who is most motivated to make sure that it's done the right way so they can contribute the most to the right clinical care.

Member Yu: And I would like to make a comment.

You know, the differences between the group facility versus the clinician, individual clinician, are very different because you put accountability and on different entities, and if you just count group or facility, basically you don't know exactly who is totally responsible, and the bottom line is to improve the practice, the, you know, the quality.

So, I felt there's a distinction between that.

And as far as for, you know, the consumers, you know, CMS has this physician comparing website, and this is a very meaningful metric for the public to look at it because they can make better decisions in order to seek better care to which physician to go.

If you hold nobody accountable, just blur the group and facility, then no one would held accountable in general. That's just my thought.

Member Pollak: I don't want to bog us down, but at my own health system, I think we do something like 300-something thousand CT scans just from reaching out to people here.

At the individual level, reading in, as you say, an outpatient setting, a lot of times -- and this is what plagues individual accountability metrics -- you just don't get those numbers.

So even though you have enough, that there might be some level of reliability, it's really hard to distinguish between good and bad, which is a huge

problem with individual accountability metrics.

So, I don't know, I'm a little challenged to see how -- I don't know, I guess I'm just repeating what other people have said.

If I had to pick one, I wouldn't pick it at the individual level in this setting, I'd pick -- unless you have data that shows that it really worked in the outpatient setting, drive improvement.

Ms. Bal: This is Poonam from NQF. Just bringing us back together on this discussion.

We've heard a few recommendations. There's been some recommendations to only have the facility level.

There's been some recommendations about combining together individual and group clinician.

There's been another recommendation about potentially combining these measures so there's only one.

I think there's a lot of options here. I think, you know, Rebecca and her team have provided some clarity on, you know, why they made certain decisions.

I think at this point unless there's additional feedback, we can just recommend that Rebecca go back to her team and just discuss do any of these suggestions make sense, and can we find a way to just better align these measures and harmonize them even more, if possible?

And I'll pause to see. I think someone was trying to speak up.

Co-Chair James: Can I make a comment right quick?

I'm kind of aligned with that in the sense that I think what we need to do -- these are use and usability issues, it seems to me, and we don't have a lot of data in that area, and in the next three years when

these three come back, we'll have a lot more data and we can make a lot smarter decision about whether or not to combine these, and of course the proponent of the measures will have an opinion, but that doesn't mean we have to go with it.

But we need more data to see how this plays out, and who is going to use it and how, in my opinion.

Ms. Bal: That's a great point, John. Any other thoughts before we move forward?

Co-Chair Yealy: Oh, I think that plan, we've shared what some of the potential pathways are, and I don't think there's anything we vote on specifically. Is that correct?

Ms. Bal: Yes, that's correct. This is more of just a discussion.

And then we'll record these recommendations and, you know, make sure they're in the report so that next time, we know that this is something that we want to discuss.

Co-Chair Yealy: Okay.

Ms. Buchanan: All right. So moving on, Sean, can you take us to the slide for -- oh, a little too far.

Co-Chair Yealy: Does that talk about the pediatric measure versus the --

(Simultaneous speaking.)

Ms. Buchanan: Yes.

Co-Chair Yealy: Yeah? So that's an overlapping but different question about if these move forward, do we need an independent pediatric measure?

Ms. Buchanan: Right. Sean, can you go back one more? Yes, this slide.

So, we've also talked about this measure as well today, and I just want to read off for you the

developer's remarks, which I think they've also spoke to earlier today as well.

Measure 2820 was developed for the same UCSF development group as the current proposed measure that calls for imaging facilities to access their radiation doses in children against published benchmarks, and it provides a framework to improve doses exceeding benchmarks.

In contrast, the proposed new measure is specified in adults.

Measure 2820 was the first generation pediatric measure, and the new measure is a second generation adult measure that incorporates the stratification by clinical indication adjustment by a patient's size and image quality.

The UCSF team plans to update Measure 2820 in a subsequent review cycle to include stratification for clinical indication and assessment of image quality, and will reflect harmonization with the newly proposed measure.

And just as a reminder, we won't be voting on anything, we're really just looking to see if these measures as they currently are are harmonized to the extent possible, or seeing if there are any differences that impact interpretability, and add a data collection burden, and whether or not the differences are justified.

So, any thoughts around this?

Co-Chair Yealy: Again, mine would be to ask the developers about this. Is the incremental need for both the same once one or more of the 3600 series is in place?

Because it does include -- I realize there's more focus, but do we really need to have both of them together, or could it be accomplished -- the same basic quality push be accomplished with the broader

measures, which are the new ones?

Ms. Bal: Don, is that a question for the group or for the developer?

Co-Chair Yealy: For the developer. That's our recommendation for it, can they consider that?

Because in this particular case, it's the same general group developing, the steward. I realize it may not be the same exact people, that's not lost on me, but --

(Simultaneous speaking.)

Dr. Smith-Bindman: It is the same exact people as the --

Co-Chair Yealy: I didn't want to presume it because I don't have that in front of me, but --

Dr. Smith-Bindman: It's the same exact people.

So, I would say the measure that you've been discussing today would be second generation measures, and the pediatric measure is a first generation measure.

So, I think more work is needed in the pediatric realm to determine the quality thresholds, as was done in the adults.

I'm hoping and planning to do that, and update the measure.

If you're saying in the future that you could potentially, once the pediatric measure gets to version 2.0 where there's quality measurement of the images, and when there's an eCQM developed to report on it, if that happens, then I would agree with you it would be possible to integrate the adult and childhood measure.

Unfortunately, I don't currently have the resources to develop an eCQM in children comparable to the adults

overnight.

That will take considerable resources and it's definitely on my to-do list, and once that measure has been sort of updated to have a more sophisticated determination of CT category using claims, an eCQM that allows measure and quality, then in the future it could be incorporated into the adult category, albeit with different categories and with different definition out of range for excessive dose or image quality.

Co-Chair Yealy: Other thoughts from the group?

Okay. Poonam, I think we have all of our comments clarified.

Ms. Buchanan: Okay, great. So moving on to our final related measure. Thank you. So I want to provide the developer's rationale for these measures not being competing.

So, the proposed measure assesses radiation doses by clinical indication, thereby allowing consideration for the reason for imaging.

Similarly, it assesses radiation doses according to thresholds determined by the underlying clinical indication for imaging rather than to observe doses without consideration if the doses are appropriate for the underlying indication.

The proposed measures' denominator includes nearly all diagnostic CT exams in adults.

Thus, the proposed measure inherently considers the clinician's subjective choice of imaging protocol, which is the single most important predictor of radiation dose.

And the proposed measure includes assessment of image quality as a means of protecting the diagnostic value of CT imaging from the unintended consequences of excessive radiation dose reduction.

I believe we did receive a lot of thoughts around this, so I'll open it back up to the committee now.

NQF Member and Public Comment

Co-Chair Yealy: Others' comments? I don't personally have anything more to add. The silence is deafening.

Ms. Buchanan: Yeah. I guess if there are no other thoughts or suggestions, we can keep it moving. All right.

So Sean, if you could go -- so I will -- yeah, I think I'm doing this. So the floor is now open if any members of the public or NQF members would like to provide a public comment.

I'm seeing Karen Campos. You would like to provide a comment?

Ms. Campos: Hi, can everybody hear me okay?

Ms. Buchanan: Yes.

Co-Chair Yealy: Yes.

Ms. Campos: All right. Great. Well thank you for allowing me to make this comment on behalf of the American College of Radiology.

The ACR has stated specific concerns through NQF's pre-evaluation public comment that have been incorporated into the materials for review in today's meeting.

These comments were primarily regarding aspects of the acceptability of these measures. While the measure developer did provide some responses, our concerns about measure validity, feasibility, and usability remain unaddressed.

For example, the measure submission mentions calculation of patient size using images and global noise calculation, which are unstructured and not

standardized processes.

Even if these feed into structured data elements, there appear to be underlying, unstructured components that warrant further validation prior to multi-site or broad use of these measures, particularly as part of payment or accountability programs.

The ACR's longstanding ongoing membership with NQF has been based on our alignment with NQF's mission and goals as a multi-stakeholder organization with principles around consensus, transparency, and collaboration through a scientific measure evaluation process.

While numerous colleagues, organizations, and the ACR agree on the importance of measure concepts of radiation dose optimization and appropriate imaging, we remain concerned whether these measures, if implemented, will produce consistent and credible results about the quality and safety of radiological care, whether the measure users will understand the results and find them useful for quality improvement and for decision-making, and whether the measures data can readily be available for measurement without undue burden.

Thank you for the opportunity to state these concerns. We'll provide additional detail through the NQF public comment forum.

Ms. Buchanan: Thank you, Karen. We will ensure that is reflected in our technical report. Dr. Mahesh, you have a comment you'd like to make?

Dr. Mahesh: Yes. Thank you. My name is Dr. Mahadevappa Mahesh.

My practice, I'm a medical physicist and I'm talking on behalf of the AAPM, American Association of Physicists in Medicine.

AAPM is a primary organization of medical physicists

in the U.S. We represent more than 10,000 medical physicists. And we believe the proposed electronic quality measures -- yeah, I want to thank the developer for bringing all these things, but there are a lot of limitations with these measures, and especially with regard to developing a consensus document, scientifically consensus it's not, and also the measures seem to be misleading.

They're not based on a community and scientific consensus, and simply are going to add an administrative burden while potentially undermining the value and safety of the imaging per patient.

Because image quality is such a moving target and we are as medical physicists working with the radiologists.

We are working on developing a consensus image quality metrics.

It's not an easy way, but creating something like a global noise, which is very arbitrary and also developing these clinical indications based on one group consensus cannot be acceptable to everyone, so it's going to be an undue burden for somebody to use this one.

In addition, this is done by a particular one group, which is not usually accepted.

In fact, radiation protection is important. Radiation dose, limitation and optimization is important.

In fact, the NCRP report, which came out recently, has shown in the last ten years that doses are going down, but there's work to be done. But not the way it's been set up here.

One is for example, on feasibility of collecting and reducing measure burden, there seem to be a lot of inadequate addressing the complexity of the CT categorization.

I understand the developer had developed a number

of clinical indication, but they are limited and they're also overlapping, and basically it's developed by one group, and it's not universally accepted how the indications.

Again, image quality is a very tough aspect to be imposed on somebody, so that what happens is like, if this measure is there, it can drive some of the clinical people just to go down on the radiation dose, not worrying about the image quality that can impact the patient outcome.

The other thing is like, regarding the scientific acceptance, AAPM also strongly feel that it's misleading the representation of image quality without any scientific verification because images will be judged to be acceptable without any relatable (phonetic) measure, reproducible measure, which the physicists, we feel strongly this has a limitation.

The other thing is like, the uncertain quality in patient size estimation, which is anchored based on BMI. It does not include a presentation of the patients and also varying habits (phonetic).

The other thing is like regarding the limited expertise in the -- regarding the CT technology, and image quality there are a lot of scientific issues which we are concerned about.

As the physicists in the AAPM has done -- has been doing a lot of work in this regard, they're trying to establish some standard protocols.

Even though there's no standard protocol, we already have some way of observing on a broader level.

For example, a site undergoing an accreditation program has to submit the protocols, which has to meet certain radiation dose level.

That to a large extent has cleaned up a lot of this clinic (phonetic), and I do agree with the developer there is a variability among the institution, and

among the clinic there are some variability, but this measure will not solve that question. In fact, it's going to complicate and become more onerous on the user.

So, respectfully from the AAPM, I'm speaking on behalf of the AAPM, telling these measures are not a good measure to be accepted by the NQF, especially since it's not a consensus-driven measure, it's more like driven by one particular group, and if a site has to do these things, they have to depend on this particular new company.

Again, I was heartened to hear that they're going to provide a way of doing it free for everybody, but it's not that easy for a company to provide some XYZ hospital to develop their own -- develop this method free of cost.

So, overall we strongly feel that these measures should not be accepted at this point of time. Thank you.

Ms. Buchanan: Thank you, Dr. Mahesh. And we have a final public comment from Francesco Ria. We're already over time, so please keep your comment concise and to two minutes.

Dr. Ria: Absolutely. I'll try to be in that two minutes.

Ms. Buchanan: Thank you.

Dr. Ria: Everybody, thank you. It was a great discussion. We follow the whole discussion today. It was very interesting.

Allow me to start with a personal experience. Last week I had a pain in my back, I called my doctor, he gave me ibuprofen, and he said it is an anti-inflammatory drug, it can help with pain relief.

She didn't tell me that it's a drug that can cause bleeding in my stomach and stroke. What I'm trying to say is that we all evaluate medical procedure based on risk to benefit balance.

And that we cannot announce one, disregarding the other one. This discussion about radiation risk and harm in -- this discussion about evaluating the quality of a radiological procedure, of a medical procedure only evaluating the risk, we think it's very dangerous, and it is unique in the medical field.

I work with ray lat (phonetic), Duke University is one of the leading laboratory devoted to the development of quality and safety measures in medical imaging, and we have published extensively on the subject with over 30 papers in the same areas that we're discussing today.

We absolutely recognize the importance of those quality measures. That's the focus of our work.

However, we believe that the proposed measures fall short of scientific rigor, verifiability, and transparency.

In particular, we believe that there is an inadequate assessment of radiation burden. Some of the metrics that have been proposed are not based on any consensus of the community.

There is an incomplete and an inaccurate representation of image quality.

Any assessment of image quality needs to be anchored to full disclosure of the metals (phonetic) and of the consensus of the community, and this is not the case here.

We are neglecting the impact of the image rendition and variations. And most importantly, we believe that the measures are not ready to diagnose in performance. In example, we are not considering at all the detection.

Subjective preference of the radiologist is not image quality. We need objective metrics and we need objective methods to assess the quality of an image procedure.

We believe that there is also an unverified assessment of patient size, we talk about weight, but I mean, there are different way to calculate weight and a patient size and BMI, and we also believe that there is a lack of guidance towards compliance, and there is a lack of support from the manufacturers.

That's why we cannot support the proposed measure. Thank you.

Ms. Buchanan: Thank you. Thank you all so much for providing your public comment. Now we're going to -- next slide, please -- move on to next steps.

So Hannah, can you give a quick overview of our next steps?

Next Steps

Ms. Ingber: Yes, thank you.

So, just quickly to remind everyone of the -- the next steps in our consensus development process, staff will develop a draft report detailing the committee's discussion today and all the recommendations that you have made about the measures.

This report will be released for a 30 day public and member comment period, and then after that period ends, staff will compile all the comments received into a comment table, which will then be shared with both developers and the committee members.

On the post-comment call, the committee will reconvene to discuss those comments, and staff will incorporate those comments and response from the committee into the draft report in preparation for the Consensus Standards Approval Committee's meeting.

The CSAC will then meet to discuss the committee's recommendations for endorsement, and then endorse the measures.

All measures that are endorsed have an opportunity

for the public to appeal that endorsement decision, and measures that are not endorsed do not go through appeals.

Next slide, please. We were prepared to have a day two but we got through all the measures today. Thank you everyone. So we don't need our meeting tomorrow.

But the draft report commenting period will last from March 30 to April 26. The committee post-comment web meeting is scheduled for June 3, and the CSAC review will take place in late July 2022.

The appeals period for those endorsed measures will last from July to August 2022.

Next slide, please. Regarding spring 2022, which is the next cycle, the intent to submit deadline was January 5 and we did get five new measures and one maintenance measure for the Patient Safety Committee.

One complex measure has been sent to the SMP for review, and we'll be discussing those and sending meeting invites out at a later date. Next slide, please.

As always, feel free to contact us at patientsafety@qualityforum.org if you have any questions, and feel free to visit our project page for that report that I just mentioned in public commenting options.

I'll hand it back to you, Erin, for closing.

Ms. Buchanan: Thanks Hannah. So, on behalf of our entire team here at NQF, I'd like to thank our co-chairs for leading us through a very rigorous discussion and helping us get through every single measure today.

Everything was great. Thank you for your patience and thank you for staying with us after 5:00 p.m., and now I'll turn it over to Don and John to let you have any last words.

Co-Chair Yealy: So I thank the NQF staff for putting together not only incredible preparation materials, but helping run the meeting and run the organization.

I thank the developers for the hard work and the clear responses today, and openness to feedback. I thank the people who signed on to give us public comments, and I also thank all of the other committee members and John. I know how much work this is on a volunteer basis.

It's why we wanted to be both rigorous and respectful of your time. I really appreciate it.

I know that everybody's professional lives are full and their personal lives even more-so, and we want to be an important and positive part of that, but not a drag.

So thank you very much, and we'll be back in touch.

Ms. Buchanan: Thank you all. Oh, John, you're on mute.

Co-Chair James: Okay. So yeah, I echo Don's comments. This is a really magnificent team.

The expertise from the people in the field, the NQF people. We worked well together today and we got this done sooner than we might have otherwise done.

So thank you for the cooperation in moving forward. I'm reminded that perfection is the enemy of getting stuff done.

So sometimes we have to settle for less than perfect to get things moving, and so thank you for all your talent, your time, and your wisdom.

Adjourn

Ms. Buchanan: Thanks. Have a good evening, everyone.

(Whereupon, the above-entitled matter went off the record at 5:09 p.m.)