NATIONAL QUALITY FORUM
+ + + + +
GASTROINTESTINAL/GENITOURINARY ENDORSEMENT
MAINTENANCE STEERING COMMITTEE MEETING
+ + + + +
MONDAY
AUGUST 27, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Andrew Baskin, MD, and Christopher Saigal, MD, Co-Chairs, presiding.

PRESENT:

ANDREW BASKIN, MD, Aetna, Co-Chair
CHRISTOPHER SAIGAL, MD, UCLA Medical Center, Co-Chair
LILIANA BORDEIANOU, MD, Massachusetts General Hospital

ZAHID BUTT, MD, Medisolv, Inc.
ROBERT ELLIS, Consumers' Checkbook
NANCY FALLER, RN, MSN, PhD, CWOCN, Nursing for Wellness
ED GILL, MD, Virginia Commonwealth University Medical Center
JOHANNES KOCH, MD, Virginia Mason Medical Center
JENIFER LIGHTDALE, MD, MPH, Children's Hospital Boston
ALAYNE MARKLAND, DO, MSc, University of Alabama at Birmingham
PAUL MERGUERIAN, MD, MS, Seattle Children's Hospital

JOHN MORTON, MD, MPH, Stanford University
ANNE PELLETIER-CAMERON, MD, University of Michigan Hospitals & Health Centers

STUART REYNOLDS, MD, MPH, Vanderbilt University Medical Center

PHILIP SCHOENFELD, MD, VA Ann Arbor Medical Center

JUDITH TOBIN, PT, MBA, Centers for Medicare & Medicaid Services

NQF STAFF:

TAROON AMIN, MA, MPH

HEIDI BOSSLEY, MSN, MBA

HELEN BURSTIN, MD, MPH

ANN HAMMERSMITH, JD

KAREN PACE, PhD, MSN

ASHLIE WILBON, RN, MPH

EVAN WILLIAMSON, MS, MPH

C-O-N-T-E-N-T-S

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

9:01 a.m.

MS. BOSSLEY: Hi. I am Heidi Bossley. I am the Vice President of Performance Measures.

So, I get to welcome all of you here today. We really appreciate you participating.

I know staff have taken you through this, but, again, you are a unique group because you are piloting-out a possible redesign of how we go about our consensus-development process. So, you are going to see quite a bit of active participation from the staff as well as at the end of the session tomorrow, hopefully, you will have some time to evaluate how this experience to date has been. And so, we are very excited to have you here.

I am going to turn it over to Andy and Chris.

CO-CHAIR SAIGAL: Yes, I am Chris

Saigal. I am a urologist from UCLA.

Andy and I are the Chairs here. Our job is to make sure the trains run on time and you guys get out of here on schedule, and to facilitate an open discussion. A lot of preliminary votes show some differences of opinion. So, it is a systematic process. We will walk through it, and I am sure it will be interesting.

CO-CHAIR BASKIN: So, good morning, everybody.

My name is Andy Baskin. I am a rookie at this particular Steering-Committee-type event. So, please bear with me. Chris will be the senior director here.

(Laughter.)

But I have been on NQF's Consensus Standards Approval Committee for the past about year and a quarter. So, that is my big credential here and how I got to this Steering Committee.

And just I get the embarrassment

out of the way really early, I did forget to pack a belt. Now it's all out and we're fine. And hopefully, today I will find one somewhere in a store to buy.

(Laughter.)

But, anyway, welcome. I am looking forward to this.

MS. WILBON: Actually, I guess Taroon and I should also introduce ourselves.

My name is Ashlie Wilbon. I think I have emailed with everyone here, and you have gotten lots of emails from me. So, here is the face behind the emails.

Thank you, everyone, for coming. We are really excited to have you all here and get started. By all means, as Heidi mentioned, we are really looking forward to getting your feedback on how things went today, and you will be somewhat guinea pigs. We are trying some new things in several areas.

So, thanks and welcome.

MR. AMIN: And my name is Taroon Amin.

It is a pleasure to see all of you in person, and I am hoping we will have a very successful and productive two days.

I don't think I have anything else to add. Maybe just introduce Evan as well, since I stole his seat.

(Laughter.)

MR. WILLIAMSON: Evan Williamson. I am the Project Analyst. I am sure you have all seen a lot of emails from me. I am happy to be here. I am looking forward to the day.

Thank you.

MS. WILBON: We are going to have Ann Hammersmith, who is our General Counsel, walk you guys through the introductions and the disclosure-of-interest process. And we will go through that.

Thanks, Ann.

MS. HAMMERSMITH: Good morning, everyone.

As Ashlie said, we are going combine introductions with the disclosure of interest. Probably several months ago, you all received a rather lengthy form from us, and we asked you to fill it out and tell us about yourself and your activities.

What we do at the first public meeting of every committee is we go around and do an oral disclosure, just go around the table. We don't expect you to recount your CV. In fact, please don't because we will be here all day. We know that you are experts, and that is why you are on the Committee.

What we do ask you to do is to reveal anything to the Committee that you think is pertinent to your service on the Committee and what is before the Committee. We are especially interested in your revealing grants, research funding, and consulting activities if they are relevant to the subject matter before the Committee.

I want to remind you of just a few

things. You serve as an individual on this Committee, not as a representative of an organization, including your employer or any organization that may have nominated you to serve on the Committee. Often, I hear Committee members innocently say, "I am Suzie Smith, and I am here representing the American Society of" fill in the blank. Actually, you are not; you are here as an individual subject matter expert.

The other thing that I want to remind you of is, because of the unique nature of the work that we do here, you could have something that should be disclosed even where no money passed hands. So, for example, if you served as a volunteer on a committee for a professional society that had something to do with the work before the Committee, that would be something that we would look for you to disclose. Just because you disclose doesn't mean you have a conflict. The idea is to be open and transparent.

So, with that, I am going to ask you to go around the room, introduce yourselves, tell us who you are with, and then let us know if you have anything to disclose.

So, would the sashless Dr. Baskin like to start?

(Laughter.)

CO-CHAIR BASKIN: So, my name is Andy Baskin. My place of employment is Aetna.

In terms of disclosures, the only thing I need to say is one is I am on NCQA's Committee for Performance Measurement, but that is not a conflict with any of this work we are doing, but just so people know.

And the other thing is that I noticed that one of the measure developers is ActiveHealth, which is a subsidiary of Aetna. I had nothing to do with those measures being developed nor wasn't even aware of them until they were submitted here. So, I don't see that as a conflict. So, I am hoping that I could still vote in that situation. Okay.

Thank you.

Other than that, nothing to disclose.

CO-CHAIR SAIGAL: Okay. So, Chris Saigal. I sit on two AUA committees. One is a Quality Improvement Patient Safety Committee, and the other is the data like Registry Committee.

Research-wise, I lead an NIH-funded project to look at quality, costs, and access in urology, and incontinence is one of the topics that we do research on.

MEMBER FALLER: I am Nancy Faller. I am an ET nurse clinical specialist from Massachusetts. I am in private practice doing consulting at this point. I am on the board of one journal that covers ostomies, wounds, tubes, continence, et cetera.

MEMBER BORDEIANOU: I am Liliana Bordeianou. I am a surgeon at Mass General Hospital. I am a colorectal surgeon. I served as a consultant on AMS. I do think

they make slings and meshes. My consulting had nothing to do with that, but here it is.

Also, I ran into somebody in the airport who is going to give a presentation here who turns out to be working at my hospital. She is a urogynecologist who I know very well.

MEMBER REYNOLDS: I am Stu Reynolds. I am a urologist from Vanderbilt in Nashville, and I don't have any disclosures.

MEMBER MARKLAND: Hi. I am Alayne Markland, and I am a geriatrician from the University of Alabama at Birmingham. I am also associated with the Birmingham VA and have a dual appointment at point.

I don't have any disclosures directly related. However, I do research in urinary incontinence and lower urinary tract symptoms, and receive funding both from NIH and the VA.

MEMBER ELLIS: I am Robert Ellis. I am the Director of Operations and Online

Information Resources for the Center for the Study of Services and Consumers' CHECKBOOK. I don't have any conflicts to disclose.

MEMBER TOBIN: Good morning.

Judy Tobin. I am a Technical Advisor for the Centers for Medicare and Medicaid Services. I am here in a non-voting capacity, evaluating the two-stage process.

MEMBER MORTON: I am John Morton from Stanford University. I am the Section Chief for Minimally-Invasive Surgery and a Director of Quality for the hospital. I chair the Surgical Champions' Forum for the American College of Surgeons.

MEMBER BUTT: I am Zahid Butt.

MEMBER MORTON: No disclosures.

MEMBER TOBIN: Sorry.

I am Zahid Butt. I wear a couple of different hats. I am a member of gastroenterology group in Maryland, a 15-person gastroenterology group. I also am CEO and CMO of Medisolv, which, amongst other

things, implements quality measures that are NQF-endorsed and other quality measures. I do not have any conflicts.

MEMBER GILL: I am Ed Gill. I am a urogynecologist at VC Medical Center, Richmond, and I have no disclosures.

MEMBER KOCH: Hi. Johannes Koch. I am a gastroenterologist from Virginia Mason in Seattle. And I don't think it is a conflict, but I do serve on the AGA Political Action Committee.

MEMBER MERGUERIAN: I am Paul Merguerian. I am the Chief of Pediatric Urology at Seattle Children's Hospital. I am also interested in continuous process improvement and chair the Quality and Safety Committee of the operating room at Seattle Children's.

I have nothing to disclose except I am an Associate Director of the Journal of Urology and, also, an Associate Director of a textbook on pediatric urology.

MEMBER LIGHTDALE: Good morning.

I am Jenifer Lightdale. I am a pediatric gastroenterologist at Boston Children's Hospital, and I don't think I have any disclosures. But I obviously do a lot of society work. I think the biggest thing is I am on the Standards of Practice Committee for the ASGE, which does write guidelines about colonoscopy and abnormal detection rate, et cetera. I am also involved in various journals, but I don't think that will really matter.

The other thing that you guys had pointed out is I occasionally do formula reviews, whether an infant should get a specialized formula for a Fallon Community Healthcare plan, and it pays very little.

(Laughter.)

MEMBER PELLETIER-CAMERON: Anne Pelletier-Cameron. I am a urologist at the University of Michigan.

No disclosures other than someone

who is presenting a measure today happens to be one of my partners that I work with, but I didn't know this until he tried to sign out his patients to me when I was coming here.

(Laughter.)

MS. HAMMERSMITH: Okay. Are there any Committee members on the phone? No? Okay.

All right. Thank you for those disclosures.

Do you have any questions of me or anything you would like to discuss with each other based on the disclosures this morning?

(No response.)

Okay. Thank you. Have a good meeting.

MS. WILBON: So, we do have one other staff member here, a very important one, who we failed to introduce, Karen Pace. I will let her introduce herself. She is sitting over here.

DR. PACE: I am Karen Pace, one of

the Senior Directors at NQF. I work with all our Task Forces and CSAC on our major evaluation criteria. So, I am interested to see how things go today.

MR. AMIN: Quick, Ashlie? Again, just noting that this is the first time we are launching a two-stage process. One of the components of this two-stage process was a technical review period in which we asked measure developers, prior to submitting concepts that you will be evaluating -- Karen Pace and Alexis Forman were leading that effort prior to the measures coming to you, the concepts coming to you. So, at various points, Karen will offer some additional insights based on the technical review process that these concepts underwent.

MS. WILBON: Just a note for myself and for you guys, too, and we have a transcript that is produced after the meeting. So, we can remind each other; we can help each other through the day to use your microphones.

That is how we kind of keep track and record the conversation that went on. So, if you see that I am talking and my light is not on, please let me know, and we will do the same for you.

So, a few logistical things that we will point, where the restrooms are, and for those of you who have not logged in yet to the internet, the wireless access log-in and password are on the screen, and we have some bullet sheets we can pass around as well, but I think everyone is in.

So, the restrooms, if you walk out of this conference room and walk towards the -- I don't even know, actually. Through the glass doors, and then make a right, yes. Okay.

So, Taroon and I are going to start out with a brief introduction with some overview slides. We are actually going to walk you guys through the evaluation of the first few concepts. Because this is a new

process, we kind of wanted to make sure that everyone is clear on how to apply the criteria, the things you should be thinking about as you are applying the criteria to make sure that everyone is on the same page. So, we will go ahead and get started with that.

MR. AMIN: And I also will just take a moment to say that we also recognize that many of the Committee members are new to NQF and also new to this process. So, we encourage you at any point, if there are any questions or any clarifying comments, please stop us because the process of evaluating each of the concepts will repeat itself over and over, over the course of today and tomorrow. So, any clarifying questions that you have, please ask them.

As Andy pointed out during one of our calls, it is helpful that you are a newer group because this is a new process. So, we are all kind of learning together in the spirit of continuous process improvement.

This is one of our tests to see if this process actually works or how it works, I should say.

So, with that, I will turn it back to Ashlie and maybe we can get started with the preliminary slides.

MS. WILBON: So, we are just going to do a few introductory slides here to get everyone on the same page here this morning. Some of these you may have seen; some of them will be a little bit different. But, again, it has been a while since we had our orientation. And so, we will just go through it again for the sake of clarity.

So, the purpose of these slides -- we have already done the disclosures -- we will do a quick project overview and scope of the project. We will, again, kind of do a quick overview of what the two-stage process, what we are proposing it will actually be and where we are in this process so far.

We will talk a little bit through

the actually meeting process. So, the steps will go through to evaluate the concepts and the order that we will go through the criteria.

And then, Taroon is going to go through and actually walk you through the first concept and each of the important subcriteria.

So, we have done that already. Those are the Steering Committee members.

So, NQF is a private, not-for-profit, voluntary consensus-standard-setting organization, public and private partnership, multi-stakeholder board of directors. We have eight stakeholder Councils, of which we try to get representation on each of the committees, so that we do have that multi-stakeholder perspective. We don't always get all of those on the committees, just because the content areas tend to be -- we try to, obviously, seat people who are experts in the content area. But those are our eight member Councils.

So, our mission, essentially, is to build consensus, endorse national consensus standards, which is the process that you guys are involved here in with the Steering Committee, and to promote the attainment of national goals through education and outreach.

So, NQF endorsement essentially allows a place for standards to get endorsed. So, there is a standardized body of measures for people to reach, or for organizations and providers to have as a toolbox to assess quality and, also, to allow kind of equal comparisons across different entities, and that they are all using the same types of standards.

The fact that our process is very regimented -- we have very standardized, specific criteria for importance, scientific acceptability, usability, and feasibility -- it has become a standard in the way that people like to use measures that have been through this process because they know that

they have been well-vetted.

So, a quick project overview here. Again, it is a two-stage process. This first stage is focused on the concepts. What the concept is is essentially the numerator, the denominator, and the details associated with each of those; the exclusions. We did ask for some information if they had some preliminary codes or what have you. We also asked for some information about if they had an idea about how they might want to risk-adjust, if it would be risk-adjusted. And we have a slide with some of the other specific components, but it is only a part, the beginning stages of an actual measure that would be specified further.

And so, the purpose of this process is to review those components against the importance criteria and the additional information they submitted for evidence, impact, and opportunity. And then, hopefully, at the end of this process, we will have some

approved concepts that will move on into stage two, in which the developers will submit the remaining information associated with those concepts. So, the actual specific codes, the algorithm associated with how the measure would be implemented, and so forth.

So, here is the project scope. This was kind of the list of conditions that we put out when we did the call for measures. Obviously, we didn't get measures in all of those condition areas, but wanted to just quickly point out, we do have a couple of pediatric docs on the Committee, of which we didn't actually get any measures specified specifically for the pediatric population. But because we tried to have endorsed measures that are cross-cutting, that cover various populations that are the broadest populations, you know, there may be some opportunity for you guys to provide input on whether or not this would actually apply to a pediatric population.

At the concept stage, it is actually really good input, where you can kind of provide some guidance to the developers on how to expand their population for the measure. So, I think that is a really unique kind of good aspect of the way we have divided this process up.

So, again, we are looking for cross-seeing inclusive or broad populations, so adults, pediatric, elderly, and the vulnerable populations, and for chronic-care- and care-coordination-focused measures.

We ended up getting 20 concepts, 10 GI and 10 GU. I will also point out that this Committee is divided. We initially started with a smaller group of GI measures that needed to go through maintenance and a small group of GU measures that needed to go through maintenance. So, we kind of, I guess, somewhat artificially put you guys together. The conditions aren't necessarily completed related. So, we are about half and half with

expertise and some people who bring a little bit of both. We realize, obviously, that everyone won't have specific expertise in every topic area, but do expect to provide your kind of medical knowledge and expertise to participate in the discussion and provide your vote.

So, we are about half divided. Of those that were submitted, about six of those twenty were actually maintenance measures. So, these are going to be concepts of measures that have been endorsed before, but they are going through the maintenance process, but they will still be evaluated as concepts for this stage of the process.

We received concept from eight developer organizations, which are listed here. Many of those are actually in the room today. So, they will probably have questions and you might have questions for them, and you can address them at the Committee's discretion, if you have questions throughout

the day.

So, this is a list of the GU concepts. I won't linger on this. You guys all have several handouts in your folder which I will go through shortly that have this, and you also should have them electronically as well. So, just a formality there.

And these are the GI concepts. And that is it.

I will just pause briefly. Before I hand it over to Taroon, I will just walk you through briefly what was in your folder. So, some of these we emailed electronically, and you may have them also on your thumb drive, if you downloaded some of the documents on there.

Obviously, we have an agenda in the front. The next handout, I believe, behind that is the staff review comments as well as the member comments that were reviewed.

So, part of the new process, one of the steps that we implemented as part of

the new process was an early member comment period. Normally, in our process we have the public and the members comment together after the Steering Committee meeting, in which they comment kind of on the Committee's deliberations and the votes, and so forth. But what we have done is kind of added an upfront member comment period where the members actually get to comment on the specifications early in the process and provide input to the Committee for them to discuss, actually, at the in-person meeting.

So, in those tables, you will see, for each concept, you will see a row at the bottom which will describe what the member comment is. So, we just ask that you look at that and consider that in your deliberations along with the staff comments.

Behind that should be the preliminary evaluations, which are compiled from everyone on the Committee that submitted evaluations. So, everyone that went into

Survey Monkey and went through that evaluation process, we have taken all of those results out and compiled them by measure into these tables. So, for those of you that are primary reviewers, you can refer to this table to help you kind of give a summary of what the votes were as you are going through your measures and presenting them to the group.

Behind that, I believe -- mine are somewhat out of order -- but I believe was a meeting quick guide, which walks you through the steps that we will kind of go through for the meeting to actually evaluate the concepts. It starts with the developer introduction, the lead discussant, and so forth.

And so, the Co-Chairs and staff will kind of help make sure that we are going through that process, but it will kind of help you, too, to kind of know what is coming up next and what that particular part of the process is actually about.

Behind that, you have a few

different tables, one of which is review assignments by concept. So, this is the same table that was in the memo that we sent out a few weeks ago telling you which concepts you were assigned to. We actually added a table behind that that is by individual. So, rather than you trying to find your name for each concept, we actually have your name and then, associated with that, each of the concepts that you were assigned to review along with which you were assigned as a lead reviewer in bold.

We will go through some of these in more detail. But we also have a table of existing endorsed measures. That table will come into play when we talk about harmonization and which of these concepts that got approved were actually similar and how they might need to be harmonized.

The last document in your folder should be a table which we will have you guys, two tables which we will have you guys refer

to, in order to help evaluate the evidence associated with each of these concepts that were submitted.

So, that was just a quick overview. We will direct you to the handouts as we think you might need them throughout the day or throughout the process. So, just kind of a quick orientation of what is in there. We realize it is a lot of paper, but we wanted to make sure that everyone had at their fingertips anything they might need throughout the day.

So, any questions? Did I miss anything?

MR. AMIN: No. It was very exhaustive.

MS. WILBON: Okay.

MR. AMIN: Thank you.

MS. WILBON: Exhausting or exhaustive? Okay.

(Laughter.)

MR. AMIN: It's good. It's good.

Okay. So, I will go through a little bit more of a detailed discussion of the two-stage process. The two-stage process follows the same process steps as the consensus-development process. So, there was a call for nominations, which meant all of you have gone through. We have identified gaps in the nominations process in which I reached out to many of you in this room. And again, I sincerely appreciate those that responded in a very short timeline with all that information. It seems like a number of people around this table fit into that category. So, I sincerely appreciate that because we were able to really have the content expertise that we need to really evaluate the measures in front of us.

We had an open call for standards, for candidate standards, that are in front of you. This was an open call for any developer that was developing measures in this clinical area.

We will review the candidate standards, which is where we are right now. And staff will take all the recommendations that you present and develop a draft report, which will go out for public and member comments. And there will be a public and member comment period where anybody from the public can provide input on the measures that are in front of you and your recommendations.

And then, there won't be voting at the concept period. And then, you will have the CSAC. There is an overarching body that ensures that we follow the CDP process, which is the CSAC which Andy is part of. The CSAC will review the recommendations, and they will go to the Board. And then, we will move into stage two.

Moving on to slide 15, I want to make sure that we are clear on what we are trying to achieve here. So, this is a consensus-development process. We felt that it is important to make sure that we are all

on the same page in what we mean by consensus. So, this means general agreement and not necessarily unanimity. All comments that are submitted by public and members are fully considered, and that each of the comments that is submitted by public and members are fully discussed and considered by the Committee.

So, what you will expect to receive as we go into public and member comment period is a commenting table, which includes all the comments that were received. And staff will provide a draft response to the Committee based on the Committee deliberations. But you will be expected to review those comments and provide any additional feedback that you feel is appropriate. And body members, the Committee, will have an opportunity to change their votes if they feel it is needed after the comment period.

So, that is what we are trying to achieve over the course of this effort and

when we go into the public and member commenting period.

So, on the next slide you will see an overview of the two-stage process here. So, what we are looking at in stage one is the measure concept and then the fully-specified measure that will move into endorsement. And concepts will be approved.

So, what we are trying to achieve in this two-stage process is really ensuring that the measurement community does not invest significant amount of resources to fully develop measures and test measures before there has been an opportunity to provide early input on the concept; i.e., are you measuring something that is important to measure? Is it important for us to move the needle in terms of quality improvement? Or it is simply a measure that we would expect to be a standard for all clinical practice?

And there are many quality measures that are important for quality

improvement, but they don't rise to the standard of being a national consensus standard for moving forward for all accountability functions.

So, that is what we are trying to achieve here. We are trialing this. And essentially, that is why you will be reviewing all the concepts against the importance criteria.

Yes?

MEMBER BUTT: So, Taroon, the second bullet in stage one where it says, "The concept numerator/denominator exclusion taxonomy," how deep do we get into those or the way the measure is constructed? In other words, do we sort of get into the details of how the measure is constructed rather than focus more on sort of is there evidence and is there a gap, and those types of things?

MR. AMIN: We don't want to get too far into the details on how it is constructed. I mean, if you look at the

numerator and the denominator and the way that it is designed doesn't make sense to you, this would be the time to discuss it. But as far as the details in the sense of code tables, I think that would be -- again, we are going to have to find a balance here because we are testing this. To a certain extent, you need to have the details in order to assess whether it is important to measure. But we also don't want to get too far into the details in terms of how they have constructed.

Is that a fair capture, Karen and Ashlie?

DR. PACE: Yes, I think that is true. We are trying to find the line here. But, obviously, definitions are very important to even understanding the concept that is intended.

So, we will be kind of seeing how this goes and also getting your feedback on it. But I think, obviously, we need to understand what they are intending to measure

and if that makes sense based on the impact, opportunity for improvement, and the evidence that is presented.

MR. AMIN: Yes, we are hoping that by focusing on the importance criteria, the three subcriteria components, that that will help drive the nature of the discussion. Again, the level of specificity, I think that is a little bit of a test. We may ask you that same question at the end of tomorrow, whether you had enough information to assess the concept.

Any other questions or comments?

Okay. So, I will just walk through this, again, at a really high level. Moving on to the next slide, you will see that the stage-one process, there was a technical process that was upfront in which we asked measure developers to submit at least one concept to the technical review period, which Karen, to my right, our lead methodologist, led.

At that point, we asked them to submit 30 days prior to the submission deadline and we provided them feedback prior to submitting the measures. There was a concept submission. We are at the Steering Committee evaluation. As I described, the draft report will go out for public comment, and there will be an adjudication of those public comments, and then moving on to the CSAC and Board.

Moving to the next slide, an overall timeline of where we are. The call for measures was at 6/4, and we expect that this information will go -- we are at the in-person meeting -- will go to public and member comment on September 14th. And it will go to CSAC for review on November 11th and board approval soon thereafter.

And then, we will move into stage two. For the concepts that are approved in this process, they will move to stage two.

Again, at a very detailed level,

we talked about the numerator and denominator statements.

And I will just move over to slide 20. Sorry, Evan.

So, what we are looking at in terms of stage two is that the measures that you recommend to move forward in stage two, the fully-specified measure, will also go through a technical review process, which will be 30 days prior to submission deadline. And I will also note here that between stage one and stage two the Committee will have an opportunity to provide a checklist of components that they would like to see prior to the measure coming back for stage two. So, at the end of each measure evaluation, we will ask the Committee if there is anything that they would like the measure developers to consider prior to seeing these measures again in the stage-two process.

The stage-two process will be mirroring, essentially, what you see here,

with the only exception that, once you go through the Steering Committee review, it will also go to a draft report, public comment, but there will actually be a member voting process. The CSAC and Board will actually endorse the measure for use in public accountability and quality improvement functions.

So, moving on to the next slide, the timeline for this, we will expect that the required technical assistance process will be completed by December 3rd, and the measure submission deadline will be December 19th. And we will meet again in February, early February, to review the second stage of the measures that were moved forward.

It will also go through, as I described, the public and member comment period, an actual vote. And we expect that this project will end around this time next year.

What will be evaluated in stage

two? Measures with full specifications and testing results. Testing results will include results on validity and reliability testing, and the remaining three criteria, which will be the scientific acceptability, the usability and use of the measure, and the feasibility of the measure as planned to be implemented.

Some key points to keep in mind on the two-stage process is that all maintenance measures, in addition to new concepts, will be required to go through stage-one and stage-two reviews if they make it through stage one. The evaluation does not change. We expect that all measures that are endorsed by NQF will still pass all the criteria. And developers will have 18 months to bring back full specification and testing results on reliability and validity. And if they don't by that point in time, they will be required to go through stage-one review again. The principle behind this is to ensure that the evidence is still valid for the measure as it

goes through stage one.

So, just at a high level, the way that we have currently constructed this is that you will go straight from stage one into stage two, as you think about the timeline. However, developers don't necessarily have to move straight from stage one to stage two. They could take a concept and then take 18 months to build it out and then bring it back for review at that point in time.

Does that make sense? It sounds good. Okay.

So, essentially, this is a pilot. Again, I just want to reiterate that. This is an iterative and first step to our process that was a result of a lean Six Sigma process improvement effort that we underwent at NQF to try to align the efforts of measure development and measure endorsement.

So that we heard from our stakeholders that they did not want to continue to invest resources in testing

measures when they were not going to pass the importance criteria. So, this is an effort to try to be responsible to that concern and to make sure that we are focusing on the most high-impact areas of healthcare improvement.

So, again, what we are testing here -- and we will look for feedback from you all at the end of day two -- is some feedback around the effectiveness of the technical review process and the revised measure submission form, the scalability -- that is what we are looking for in general from our internal process -- and, also, looking at the overall cycle time. Sorry. This is not areas that we are looking for feedback from you. This is how we are testing the effectiveness of this process in general.

In stage two, we will also be looking at the overall cycle time and the effectiveness in terms of the staging of the evaluation. So, what we will also be looking for is how many measures actually made it from

stage one to stage two and what were the reasons for not making it.

CO-CHAIR SAIGAL: Taroon, if the measure developers in this round don't submit into stage two for the February meeting, is there a mechanism for that, if they take longer to sort of test and develop the data around the measure? Is there another venue for them to resubmit the measure within 18 months.

MR. AMIN: So, the plan is that we would have something for them to submit into within 18 months. We are working on moving into standing committees in the future, and that would be the goal. The exact project and the exact time period is a little bit up in the air.

But if there is anything else you want to add, Heidi, on that, in particular, feel free.

MS. BOSSLEY: No.

MR. AMIN: Okay. All right.

By the way, I don't know if we had the opportunity -- Heidi, did you have the opportunity to introduce yourself?

MS. BOSSLEY: I did.

MR. AMIN: Oh, you did? All right. There we go. Excellent.

(Laughter.)

And Helen Burstin also joined us. Apologies that there was not enough room at the table, but she is in the back.

Anyhow, next steps: as far as where we go from the stage-two process, just make sure you have these dates listed for yourself to ensure that we are good for stage two in 2013.

But the orientation will be in early January. The Committee evaluation, the training webinars, which you all participated in, will be January 22nd and 24th. And the in-person meeting will be in early February, with a conference call to discuss the public and member comments.

So, I will turn it back to Ashlie to discuss the meeting process for today.

MS. WILBON: Some of this I have already been over. So, I am not going to take a lot of time. I think everyone is probably anxious to get to the meat of this meeting, which is actually evaluating the concepts. So, just make sure you have all the resources you need to be able to do that.

We have already been through some of your handouts. If you find you are missing anything, just let us know. We will get it for you.

Again, what we are going to do for the meeting format is we will start with a developer introduction. The way the agenda is set up is most of the concepts are grouped by developer, in addition to kind of topic area. So, we will have a developer come to the microphone, give a really introduction, two to three minutes, about what they have presented, if there are any unique attributes of their

concept that they want you to consider or any additional information.

And then, we will have the lead discussant kind of start and introduce the concept, go through and highlight some of the things that came out of the preliminary evaluations, and then open it up to the group for discussion from there.

After that, actually, the discussion of the Committee will actually align with each of the subcriteria within the importance criterion. We will start with impact, which is the criteria 1(a). We will have you guys vote. We will move on to evidence. We will have you vote. Then, we will move on to performance gap and have you vote.

In between those, both Chris and Andy are going to try to summarize the Committee's discussion, so that we are clear as a group and the staff, so that we can translate the Committee's kind of feel or, you

know, judgment or evaluation of each concept, so that it is reflective of what the actual votes were.

So, if you find during the summary that it is not quite reflective of what you heard, please, by all means, let us know and we will make sure that we have captured all of that.

Following that, we will actually have you submit a vote on your overall recommendation for whether or not you think this concept should move forward to stage two. So, we will have each of the votes on the subcriteria within importance, but, then, we are asking you to kind of, taking all of that into consideration, whether or not you think it should move forward to stage two. Again, we will have the summary of discussion.

There are several periods in the agenda that we will have periods for the public and the members to comment. We do have several members of the public and the

developers in the audience. So, that will be an opportunity for them to pose questions to the Committee or just make general comments for those in the room.

At the end of the discussion of each topic area, so at the end of our discussion today of all the GU concepts, we will have a broad discussion for those concepts that actually are approved on whether or not there are any issues around harmonization that need to be discussed. And we will walk you guys through that when we get to it a little bit more.

And then, the same for day two, when we have the discussion on the group of GI concepts, we will have a discussion at the end of that, that group of concepts, on whether or not there is any harmonization that needs to occur within those.

Sure.

MEMBER MORTON: I had a question a question about harmonization. So, if we see

a couple of measures just on the agenda that look pretty close together, we approve the items individually and then come back to them?

MS. WILBON: Right. So, if they make it through the -- which is why we have harmonization at the end, because we only really discuss harmonization for those that are actually approved. So, we do evaluate each one individually and then we kind of come back and look at the group of those that were approved and see what might need to be done. Yes, absolutely.

So, the next slide, I will just talk a little bit about the role of the lead discussant. I think we have said this many times, and there were several things that were sent out. But, essentially, we are just asking you to introduce the concept for the group, summarize what was submitted in the preliminary evaluations and anything additional you would like to add that you think the group should consider, and,

obviously, emphasizing any areas of concern that either you were troubled with or had difficulty discerning, or whatever, so that the whole Committee can comment on those things.

So, the electronic voting. Everyone should have a little remote thingy that they will be using to vote with. I think, with the exception of Judy who is non-voting, everyone should have one. If you don't, let me know. I think I have given everyone one.

At the time of the voting for each subcriteria, we will have slides up that will give you instructions. Actually, I think the next slide is an example. Yes.

So, it will be a slide that looks very similar to this. It will tell you what your voting options are. You will just hit the corresponding number on your remote and point it towards this thingy.

(Laughter.)

This is the electronic reception for how you vote.

So, it is anonymous. We won't really know who voted how necessarily. But that is how we will do voting.

And then, once all the votes will show up on the screen, we will read them aloud, so that everyone knows what they are. And then, we will move on to the next.

Do you want to just go back to that one slide?

Okay. Oh, go ahead.

MR. WILLIAMSON: Yes, just a clarification on this slide. You don't need to hit "Send". If you just press the number, 1 for yes, 2 for no, or 1 for high -- it will show on the slide -- it will record it on there.

And you don't need to worry about duplicative votes. You can keep pressing it and it will only record it once. Sometimes we find that we are like one vote short and we

ask everybody to press it again. Just make sure you are pointing at, I believe it is called a "dongle".

(Laughter.)

MS. WILBON: Oh, okay.

MR. WILLIAMSON: That is the technical term for it.

MS. WILBON: "Dongle" and "thingy" sound very similar.

(Laughter.)

MR. WILLIAMSON: So, just point there, press the button, and you don't need to hit "Send". It will record your vote.

MS. WILBON: So, the last slide, I believe the timer is set for 60 seconds.

MR. WILLIAMSON: Yes.

MS. WILBON: So, we will kind of let you know before we are going to hit "Start". So, kind of have an idea of how you are going to vote. You have a minute. Hopefully, by the time we have a discussion, everyone will know how they want to vote. And

then, again, we will display the results on the slide and read them aloud.

So, that is the meeting process itself. Does anyone have any questions or anything about that?

(No response.)

Okay. I am going to hand it over to Taroon, and we are actually going to --

MR. AMIN: Get started.

MS. WILBON: -- get started, yes.

MR. AMIN: Does anybody have any questions before we get started? I know everybody is very excited to get started.

(No response.)

So, the way that we will do this is, for the first two to three concepts, staff, myself in particular, will help to lead the discussion, so that everybody gets the hang of how this is supposed to work and we don't put too much pressure on the lead discussant right from the beginning.

So, what I will do at a high level

is introduce each of the subcriteria, and then we will ask for the lead discussant to provide their input, as Ashlie described. We will ask for some discussion around the preliminary evaluation and, also, to address the member comments, if they are relevant to the measure, in particular. And then, open it up to general discussion for those that were also reviewing the measure, and then to move forward on voting.

I would also keep in mind, in order for us to stay on schedule, we have about 20 minutes a measure over the course of the two days. So, while we want to have robust conversation, let's try to keep the comments and we will kind of rely on the Co-Chairs to make sure that we move along in that fashion, not to put any pressure, but that is kind of the pace that we are thinking about.

(Laughter.)

We have great staff support.

So, we will start with, we will

ask the measure developer, NCQA, to give a brief, two-to-three-minute introduction to both of their measures that will be considered in the first time spot. And again, this will be the nature of how we will ask the measure developers to give a brief introduction to all their measures that are going to be considered in their block of time.

So, we will ask NCQA to give a brief, two-to-three minute introduction, again, in the spirit of making sure that the time is two to three minutes. And that will be for Measure 0030 and Measure 0098.

Please.

DR. GIOVANNETTI: My name is Erin Giovannetti. I am with NCQA. I will let my colleagues introduce themselves.

MS. ALAYON: Hi. I'm Dawn Alayon from NCQA

MS. BARTON: I am Mary Barton from NCQA.

DR. GIOVANNETTI: Our measure is,

the first one I am presenting, is management of urinary incontinence. It is a patient-reported measure that has two parts. The first is a question which asks, for people who have self-report symptoms of urinary incontinence, have they discussed it with their provider. The intent of this first rate is to get at whether or not healthcare providers are discussing urinary incontinence with their patients.

The second question is, have you received treatment for your urinary incontinence? The intent here is that people who have urinary incontinence should receive some sort of treatment for that.

This is a measure which we feel is very important. It is up for maintenance. So, this has been a longstanding measure in our Health Outcomes Survey. It addresses what I think is a very important quality gap, which a lot of individuals, especially older individuals, which this measures targets,

individuals 65 years and older have urinary incontinence but do not receive treatment for it, either because they think it is a natural part of aging or for other reasons around embarrassment. This is a significant quality gap that we see that needs to be addressed.

The measure is based off of guidelines and evidence behind those guidelines suggesting that individuals with urinary incontinence should be screened for urinary incontinence and that they should receive treatment.

The question itself, if you look at the wording, is not prescriptive in what type of treatments, although we included all the guidelines in there for the recommendations for various types of treatment. But the question itself just says you should receive some type of treatment and lists some examples, but leaves it open.

And so, with that, I will be happy to answer questions, and maybe I can sit and

be by my notes to answer those.

MR. AMIN: What I will do at this point is walk through the criteria, just to make sure that we are all kind of understanding what we are looking at individually.

So, just as an overall, we will do an introduction of the criteria, and then, again, we will turn it over to the lead discussant to have a discussion. And then, we will go through each of the individual criteria. What I will do is I will introduce the criteria, the individual subcriteria, and then vote.

Again, this will only be for the first two concepts, and then I will kind of turn it over to Andy and Chris to take a little bit more of a lead here. This is just something that we are trying to see how this works.

So, what you are looking at for these two measures or for these two concepts

is to the extent that the measure focuses evidence base, important to make significant gains to healthcare quality and improving health outcomes for specific high-impact aspects of healthcare. That will be evaluated through high impact, the evidence that supports the measure focus, and gaps in performance.

Specifically, when we are looking at impact, we are looking to assess whether the concept addresses a specific national healthcare goal of priority or the data submitted that demonstrates the high-impact aspect of care. What you should specifically be looking for is the number of people and the percentage affected or dollar amounts, the specific relationship to the target population, and the category of impact selected.

And you will rate this in terms of high, moderate, low, and insufficient. The rating scale, on the next slide, high,

actually, what the definition for high is that it is based on the information submitted. There is a high confidence there is certainty that this criterion has been met. Moderate is that it is moderate confidence, and low is that there is low confidence. Insufficient is that there is insufficient information submitted to evaluate whether this criteria is met. In particular, blank, incomplete, not relevant, or responsive to the particular question at hand.

So, before we move to discussion, to the lead discussant, I will turn it over to Nancy to have the lead discussion conversation.

I would like to note for the group, first is that you should evaluate the information presented to you through the submission form. So, you are actually just evaluating the information in front of you. That will become very important as we talk about the evidence component.

The second is that Measure 0098 is actually being combined. There were three measures that were currently endorsed. Measure 0098 was the assessment of UI measure. Measure 0099 was the characterization of UI, and Measure 0100 was the plan of care UI, which have been combined in this measure that is in front of you. So, we will leave it there.

I will also make a note that we have a procedure here. The measure developer should really only speak when spoken to by the Committee. If there are questions that you would like to ask, that is fine, but we want to limit the back-and-forth.

The second thing I will just note is that I asked the developer to give an introduction to both measures, and that will be the format that we will be using. If there is additional conversation you would like to have about 0098 --

CO-CHAIR BASKIN: I think what

happened is when she did the introduction, she only introduced No. 0030 instead of 0098. So, I realize you asked for both, but this time around maybe we could get you to talk about 0098 as well.

MR. AMIN: Sure. That is your discretion.

CO-CHAIR BASKIN: Because it just didn't happen.

MR. AMIN: That is your discretion.

CO-CHAIR BASKIN: That just didn't happen, but in the future we will be a little clearer in our communication.

MR. AMIN: Sure.

CO-CHAIR BASKIN: So, yes, go ahead.

DR. GIOVANNETTI: I apologize. It is a lot of instructions flowing around. We are all trying to do the best we can in a new structure. So, I apologize for interrupting. I wanted to make sure things were clear.

Measure 0098 is a measure that is a provider-level measure. It is a measure that does combine three rates. This is for older women, women age 65 and older who are seen in the ambulatory care setting.

This has three rates in it. One is to screen for urinary incontinence those individuals who are then have a diagnosis of urinary incontinence to ensure that they had both characterization and a plan-of-care documented in their medical record.

So, I will leave it at that.

MR. AMIN: Okay. So, we will start with the lead discussant conversation.

MEMBER FALLER: This isn't fair; since I am the first one, I don't know what I am supposed to do.

But just from a practical perspective and a patient perspective, in reading it, I had two concerns. One, were they asking the patient whether the MD had initiated that discussion or whether they, as

the patient, had initiated that discussion, because I think that is really significant in terms of historically what we know about patients discussing incontinence with providers.

And the second concern that I had was the word "treatment". Even though they are given different treatments in what they are being asked, are they being asked, were you treated or were these discussed with you? Because those are very different. Sometimes the provider may give them the options of biofeedback, or what have you, and the patient may choose not to do any of them because they don't particularly like any of them. Or they may have only been suggested surgery, and they don't want surgery.

CO-CHAIR SAIGAL: Nancy, could I ask? So, I think it is good for the group if you could categorize your comments in terms of like the different things we are voting on, to help us understand what sort of portions you

are concerned about. So, there is importance to report it and the evidence behind that. I guess those are importance to measure and report which we are talking about right now.

MEMBER FALLER: I guess mine was even before that in terms of, what are we asking? I mean, I have no problem with the issues and the importance of them, the impact that they will have because of the numbers of people who are incontinent. But I have a concern with what we are going to be asking the people. So, I am not sure which one of those that falls under.

CO-CHAIR SAIGAL: I mean, we have to evaluate the measure as submitted. So, basically, if what you are saying is that, as submitted, you think that these are insufficiently defined to be important, then I guess that is how we would think of it. Is that what you were trying to say or?

DR. PACE: I think in the measure submission form, under the details, there is

a specific question that is on the survey. And so, for example, the discussing urinary incontinence question is, "Have you talked to your current doctor or other health provider about your leakage problem?" So, that is the exact language of the question that is being asked on the survey. And then, the other one is there as well.

But I think, getting back to the point that was just made, if we can talk about impact of urinary incontinence in terms of whether that general area is a high-impact area, and then we can move on to the evidence about this specific focus. But let's, first, just talk about the impact of that general area of condition, urinary incontinence.

CO-CHAIR BASKIN: So, I see people starting to raise their hands. So, we found it very helpful in the past, because it is hard to keep your hand up, if you will just put your card up, we will know that you want to speak, and then we will get you speaking.

Okay? But I think your hand up first. So, it would just make it easier.

MEMBER MORTON: Either raise it or just throw it out, right, one or the other?

(Laughter.)

The question I had is it might be useful just to have the criteria up, and that will help formulate some of the responses that we are all looking at.

DR. PACE: I would like to say that high impact is more than just number of persons and percentage affected. I mean, those are key ones that are generally brought up, the number of people affected, the resource involved. It could be, also, the consequence of poor quality could be of high impact. And sometimes even with a small volume of patients, it could be an extremely high impact area, the consequence of poor quality. So, I think just kind of keep in mind it is looking at it in context to the particular condition.

MEMBER MERGUERIAN: I am just wondering if another question should not be added. Even though it is in the summary of the evidence for high impact where you basically say that it affects their well-being, but there is really no question on quality of life and how bothersome that incontinence is to that woman. So, I think adding a quality-of-life question there or a bothersome index into that question, that survey would be important.

MEMBER BUTT: I think there is a question that says big problem or a small problem.

MEMBER MERGUERIAN: So, that is in 0098?

MEMBER BUTT: No, 0030.

CO-CHAIR BASKIN: Yes, it is not just a yes-or-no question, apparently. It is a none, little, some, or a lot kind of question.

Oh, I'm sorry, I should have my

microphone on.

I mean, it is rather obvious to me that this is a common problem. It affects a lot of people and a great percentage of the target population, which is over 65 women. Any concerns about that?

MEMBER BORDEIANOU: There is a big study in JAMA, I believe, and I can't remember the reference right now, that documented the rate of urinary incontinence in women at various ages. I think the rate of incontinence in women is over 25 percent by age 65. I said it is a common problem in women over 65 that has been very well-documented, and I can't remember the reference off the top of my head, but it has been published in JAMA several years ago.

CO-CHAIR BASKIN: So, any concerns here at all about the first impact 1(a)?

(No response.)

Then, let's move on to the next one.

CO-CHAIR SAIGAL: Yes, I would agree. That's fine. I agree. Go ahead.

CO-CHAIR BASKIN: Then, the next question we were to answer --

MR. AMIN: First, we vote.

CO-CHAIR BASKIN: Oh, we vote on 1(a) first?

MR. AMIN: Yes.

CO-CHAIR BASKIN: Okay.

MR. WILLIAMSON: So, we will now vote on the importance to measure and report for high impact. You may begin voting now.

(Vote taken.)

We have one not voting, right?

Fourteen? Okay. All right.

And the results, we have 14 high, zero moderate, zero low, and zero insufficient.

MR. AMIN: Okay. So, thank you for that.

So, high impact, you know, generally, there isn't as much probably

discussion. Maybe I should have prefaced that by saying there is probably not much debate there.

So, moving on to the next criteria that we will evaluate, the first is -- I will just wait for Evan to transition the slide.

What you are going to be looking at here is assess the evidence that supports the measure. If it is a health outcome, we are only looking at a rationale. And again, you want to look at the evidence based on the information that is submitted.

Looking at the preliminary reviews, I noticed that there were a number of comments that the Committee members made about the evidence that exists, but not necessarily the evidence that was submitted. Again, I am not referring to this measure in particular; I am just referring broadly, that there was evidence that exists but not submitted in the measure. You need to evaluate the measure that is submitted in front of you and the

information that is submitted in front of you.

And I will give you some caveats as we go through. I will also say that this is the subcriteria that will probably require the most discussion and the most clarification upfront. So, I just want to make sure everybody is aware.

So, what we are looking for is for the measure developers to summarize the body of the evidence. That would be through three different options. If the measure is based on the guidelines, to look to see the evidence that supports the guidelines under evaluation. The second is a summary of the evidence through a systematic review or, third, that the developer does a review of the evidence themselves.

There is a question?

MEMBER LIGHTDALE: I have a question. I have a question about what you mean by evidence. Is it evidence that the problem exists or is it evidence that

treatment would make a difference, or whatever is being proposed as you numerator would make a difference?

CO-CHAIR BASKIN: Right. It is evidence that whatever process or whatever is being measured would make a difference.

MEMBER LIGHTDALE: Okay.

MR. AMIN: And again, I think that was another point that I wanted to bring up, in that that seemed to have come up also in the preliminary evaluations. That is information that supports the performance gap question around the problem exists, but really what we are looking for is that this measure focus actually improves patient outcomes. And that is particularly the type of evidence that we are looking for.

CO-CHAIR SAIGAL: And so, basically, we are looking for evidence that measuring this would reduce incontinence in women. That is the idea.

Go ahead.

MEMBER BUTT: Yes, I have a question. A measure like this, where it seems to be it is submitted as a process measure, but incontinence itself is an outcome, isn't it? And it is a combination process and outcome. So, the treatment part is the process, and the incontinence itself is the outcome. So, it is more we focus on the treatment? Because it is a two-part measure, right?

CO-CHAIR SAIGAL: The resolution of incontinence will be the outcome.

MEMBER BUTT: Okay.

MEMBER MORTON: Well, the process is asking about it, isn't it? The process is asking about it?

DR. PACE: This is actually two processes. One is whether the person answering this survey has had a discussion with their care provider, and the second part is about whether they received treatment.

MEMBER BUTT: Okay. Got it.

DR. PACE: So, you are right that a measure of actual incontinence before and after would be an outcome measure, and that is what we are asking for. It is what the evidence about these particular processes in terms of achieving that outcome of resolution?

The other thing that I will just mention about this to kind of consider -- and it is a distinction between this one and the next measure -- is this is actually from the patient's perspective in terms of, did they actually have a discussion; did they receive treatment? And this would be an experience with care, and people would think of this as the patient being the authoritative source of whether they actually had a discussion and actually received treatment.

MEMBER ELLIS: Hopefully, I am not being too redundant, but just to clarify this. It struck me that the first part of this measure seemed much more focused on patient engagement, which I think it is easier to kind

of wrap our brains around its importance because we have addressed the importance of patient engagement in a lot of other measures.

And then, you have this kind of dividing line and you have the outcome piece. I am wondering how the evidence evaluation plays in those two component parts.

CO-CHAIR SAIGAL: Well, my understanding of this is that there was an NQF Evidence Committee that emphasized the importance of process/outcomes links in these measures. We can measure a lot of processes of care that may not affect what patients care about. So, they would like us to evaluate whether we, based no the evidence submitted, can find a link there. I think for many of these measures it is very hard to find a specific study that shows that process/outcome link.

There is an exception that we have and the option to use if we feel that the evidence isn't there, but that the benefits

outweigh the harms of proceeding with it. But it is supposed to be an exception and not a routine thing that we do to override a lack of evidence behind any one specific measure. So, the bar is pretty high, I would say.

MR. AMIN: So, before I move on -- and maybe, Karen, also you want to jump in here -- but for each of the components, when we have multiple components, there needs to be evidence supporting each piece. So, that is really the way you want to think about this. Again, not necessarily just related to these measures broadly.

And let me just, if I can, go through the rest of this related to evidence, just to make sure everybody understands the options in front of them. So, what you are looking for is, when you are looking at that body of evidence, you are looking to evaluate the quantity, quality, and the consistency of the evidence that supports the measure focus. We would expect that, for consistency, it is

rated moderate or high, and that the quality and quantity is rated moderate or high.

So, just moving on to the next slide, again, what we are looking for in terms of quantity, the total number of studies, not just the articles or papers.

For quality, the certainty of the evidence -- or sorry -- the certainty or confidence in the point estimates.

And the consistency is we are looking for the stability and the magnitude and direction of the clinically- and practically-meaningful benefits and harms to the patients.

So, I will go through each of these individually in terms of how you vote.

Quantity, so moving on to the next slide, you will see that we are looking for high. There are five or more studies. Two to four is moderate, and one would be low. Insufficient is that the information that was presented to you, there is no empirical

evidence or that the selected studies are from a larger body of evidence.

MEMBER BUTT: Taroon, in this context, do you consider a guideline a study?

MR. AMIN: No.

MEMBER BUTT: That is a paper?

MR. AMIN: It is not either. The guideline needs to demonstrate, needs to catalog the quality, quantity, and consistency of the evidence that supports the guideline itself. And we will go into a little bit more on that because I think that will be the nature of the discussion broadly, again, not particular to these measures, but just broadly.

Karen?

DR. PACE: There's more slides.

MR. AMIN: Yes, there's more slides.

DR. PACE: Okay.

MR. AMIN: Oh, yes.

DR. PACE: Why don't you finish?

MR. AMIN: Okay. All right. So, for quality, we are looking for high quality; we are looking for randomized controlled trials with direct evidence specific to the measure focus. We are not looking for just information related to the condition broadly. We are looking for really what the measure is specifically focused on measuring.

Adequate size for precision in the point estimates, moderate is non-RCTs with the control for the confounders, and low looks at RCTs with serious biases. And again, insufficient is that there is no empirical evidence.

And third, and probably most significantly, looking at consistency, which is often lacking in the information presented in the guidelines broadly, but that is tangential, is that for high, you are looking for estimates that there is clinically- and practically-meaningful benefits in the evidence base that is actually presented, and

similar magnitude across the studies.

And moderate would be that there's estimates of benefits and harms, but they may differ in terms of magnitude. And low is that there is estimates of benefits and harms, but they differ in terms of magnitude and benefit, and there's wide confidence intervals.

CO-CHAIR SAIGAL: So, Taroon, we are being asked to evaluate the impact of this, but also its relationship to outcomes. So, those are two types of evaluations because there is, generally speaking, a lot more evidence of the impact of the problem on people than there is about process outcome. So, how should we define that or decide on that in terms of doing this rating?

MR. AMIN: I think, again, we are looking for this particular subcriterion -- and again, Karen, if you want to add anything here -- we are looking at, in particular, for this subcriteria the evidence that supports the measure focus.

So, impact is 1(a), which we are looking for information that supports the impact of the measure focus. But here what we are looking at is really the evidence that supports the measure focus in relation to the outcome, not a cataloguing of the nature of the problem that exists.

DR. PACE: Right. So, I think your vote already showed that all of you thought that this was a high-impact area. So, now we are moving on to, what is the evidence for these processes in terms of relationship to the outcomes. And then, the third component that we will eventually get to is performance gap.

So, just to get back to guidelines, some are just based on expert opinion, some guideline recommendations and the evidence task where it is really focused on empirical evidence, but we do have an exception if it is really deemed necessary by the Steering Committee that it still warrants

having a performance measure.

And I don't know if you were going to talk about those two things.

MR. AMIN: Yes, I am going to go over it. Yes, I will go into it.

DR. PACE: But, the Evidence Task Force did make a specific distinction between guideline and evidence. As you all know, not all guidelines are created equal in terms of the processes that are used. So, that is where the Evidence Task Force really asked for a description of the summary of that evidence review, so that we could really be transparent about the evidence.

CO-CHAIR SAIGAL: Okay. So, for this group, we have evidence in this measure about importance, but nothing was submitted specifically about the treatment and outcome relationship. But the urologists in the room at least -- and some of the gynecologists -- are aware of evidence that suggests that treatment of incontinence is effective. So,

we could basically talk about that in this meeting or we could have the developer to include those data when they go to phase 2, studies showing that, you know, Kegels or surgery are effective in treating incontinence.

MR. AMIN: I think we will actually probably do both. And so, I will talk about the options for the evidence in a second. Chris, that was a great summary, and I might need you to do that again as we actually go through this.

But let me just finish this real quick, and then we will open up to questions because I want to make sure everybody is on the same page.

So, Evan, if you can go to the next slide?

What you are ultimately going to actually rate is the last column here, which is whether it passes this criterion on evidence. And we talked about the fact that

you will have a discussion around the quantity, quality, and the consistency of what is presented in front of you. And again, the only way that this criterion really passes, as you can see really, that if it is moderate or high across quality, quantity, consistency.

Now, just moving on to the next slide, what we will ask you to do is vote on the information that is presented in front of you. So, if you do not believe that the information presented in front of you adequately captures, gives you the actual information related to quantity, quality, consistency for you to adequately rate it, we ask you to first make that very transparent.

Again, one of the key considerations from the Evidence Task Force which was, again, a consensus-based group that asked how we should be endorsing national consensus standards, was that you need to make it very transparent when there is not evidence that is supporting the measure focus in what

is submitted in the evaluation, in the measure submission form.

So, you will first rate, and you will have three options. The first is that the evidence, the body of evidence, meets our guidance on quality, quantity, consistency. The second is that the evidence does not meet the guidance for quality, quantity, consistency, including that there is no empirical evidence that exists. And if there is no evidence that exists at all to support the measure focus, you have the option of invoking an exception that there is exceptional and compelling reason that the measure should be considered forward; specifically, that the benefits outweigh the harms.

And we will ask that, once you actually vote -- so, you will vote no, and then we will ask for the Committee to invoke an exception. And there will need to be a rationale that supports that exception. So,

specifically describing the body of evidence that supports this measure focus that may not have been submitted as part of the measure evaluation.

The last option you have is that there is insufficient information that was submitted to rate quantity, quality, and consistency in terms of that there is information that exists, but the measure developer did not provide it. And there, we will ask whether there is general agreement that the quality, quantity, and consistency of the body of evidence would meet the NQF criteria. And then, we would ask for a discussion and rationale.

So, if the information doesn't exist at all, that would be No. 2. And if the information exists, but was not submitted as part of the measure submission, that would be No. 3. Is that clear?

And then, as you vote no, there will be an option for you to have a secondary

vote to continue to move the concept forward, with the caveat being that the information that was presented would not meet the guidance for the NQF -- would not meet the NQF criteria as submitted. Is that clear?

MEMBER BUTT: The four different methods of grading that are given as examples, are those the only ones? Or, for example, if there is a grading system that is quoted that is not listed here, how do you sort of tackle that?

DR. PACE: That is a good question. That really gets at the heart of the Evidence Task Force recommendations that we ask for a summary of the quantity, quality, and consistency of the body of evidence, because across guidelines they use so many different grading and rating systems, that there is no way to do a crosswalk among those.

You know, you can see guidelines where grade A means it is a strong consensus opinion versus other guidelines where grade A

means there are multiple RCTs. And so, that was really part, you know, very clear as the Evidence Task Force looked at current practices and guideline development and, also, at the same time our Evidence Task Force was working. The IOM actually had two projects going on related to this same issue, one in systematic reviews and one on developing guidelines.

And so, in order to be transparent, the Evidence Task Force said we really need to have at least some summary information about the quantity, quality, and consistency of that body of evidence because we don't really know what those processes were and what those gradings -- you know, it is hard to equate those grading systems across guideline developers.

But I will say that one of the things that the developers come up against is about the status of guideline development, that guideline developers maybe have not been

very transparent around their systematic reviews or making their evidence tables available.

So, this is an area, again, where there is a lot of variation and, hopefully, evolution to higher standards. That is why we have you, as experts, around the table, to also provide guidance in this area.

We really do expect the measure developers to understand the evidence on which they are proposing National Performance Standards. But, again, there are some limitations of what is available.

So, sometimes if it is not real evident with the guideline, perhaps the guideline is based on a systematic review where that information would be available. Or there may be something through the Cochrane Collaboration or AHRQ or some other reference. But, again, that is where you all may have more information about that.

MEMBER BUTT: So, I am still not

sure if I understand. Should we sort of accept the way it is presented, that (a) means good scientific evidence was reviewed, even though it is not a specific methodology that we are familiar with? Or do we combine that with our sort of other information and knowledge and say, yes, it sounds like it is okay?

DR. PACE: So, I don't know, if you look at the measure submission, the attachment about the evidence, we ask about the guideline and the grade. And the section where we really ask for them to summarize the quantity, quality, and consistency starts at 1(c)(8) or that section on findings from the systematic review.

And so, if the information has been provided there that you can use that information and the findings from the systematic review beginning with item 1(c)(8), then you can just apply the rating scale. If that hasn't been provided in a way that you

can actually provide the rating scale, then we would say that you really have insufficient information on which to apply the rating scale and you would vote it that way.

But, then, the next question is, are you aware of the evidence that would meet these criteria or whether you think it is really an exceptional situation that we really need a performance measure in this area.

So, I will stop there.

CO-CHAIR SAIGAL: Okay. So, should we vote?

Oh, one more?

MEMBER REYNOLDS: I just want to make sure because I know that this is the first one, and we are just trying to get our hands on it. I think you are giving a good overall, but you are trying hard not to push us a little bit and point out specifics.

So, I just want make sure on this one that people do see the specifics, that if you look at the form that you can get off the

website and on the pilot submission form, I think the measure developers have done a pretty good job of giving evidence and summary. And so, they have identified two guidelines that speak to both of these, the two components of this measure. And so, they give the specific results of the guidelines with the grade of the recommendation, and then information on the levels of evidence.

So, I would point to Section 1(c)(4).4, and they specifically say that, in terms of the first part, "Health professionals should be vigilant and adopt a proactive approach in consultations with patients who are greatest risk." They gave that a recommendation grade of B. And the levels of evidence to form that recommendation are 2-plus and 3.

Well, then, they go on to the next section to describe what those recommendations are. A grade D is a body of evidence including -- you will be able to read it for

yourself. But, essentially, they are saying that the guidelines show it is the summary of the evidence on it and it is based on high-quality, systematic reviews of case control or cohort studies.

But, then, the second part, which is that we know that there is evidence that treatment of incontinence is strong -- and Chris has pointed to the point that you made -- that also is there in both guidelines. And certainly, the ACOG guideline gives those treatments a grade recommendation of A, which is based on good and consistent results.

And so, I would argue that, for the summary data submitted here, that there is enough information here that I think you could probably formulate a pretty good vote on what there is, but I don't think we have really pointed that out yet. We have been talking more about general aspects, but looking at this specific one.

MEMBER MERGUERIAN: I just wanted

to make a clarification. So, the first concept is for males and females. So, it is all people above age 65, and the second concept is just women, right?

CO-CHAIR SAIGAL: Okay. So, now shall we vote?

MR. AMIN: I think as we presented all of the information to the group, I think we still want to have the discussion on what is in front of you. I mean, if you feel comfortable that there has been sufficient discussion, it is a question, I think, out. We still haven't had a discussion from the lead discussant. If you feel comfortable that you can vote at this point, it seems like additional discussion would probably still be needed on what is in front of you, on the quality, quantity, consistency.

CO-CHAIR SAIGAL: So, Nancy, I'm sorry.

Stu, so you were a reviewer for this, right?

MEMBER REYNOLDS: I was not technically.

CO-CHAIR SAIGAL: You were not?

MEMBER REYNOLDS: No.

CO-CHAIR SAIGAL: Who else was? Anne, were you one?

MEMBER PELLETIER-CAMERON: No.

CO-CHAIR SAIGAL: No? Okay. I was I think.

So, you want to go through the whole formal process again? We already voted on importance.

MR. AMIN: We did importance. Now you have evidence.

DR. PACE: So, we did impact. This is all part of importance.

MR. AMIN: Right. Sorry. We did impact. Now we are doing evidence. Now we are doing evidence. And so, the goal here would be to have a discussion around -- I introduced what the quantity, quality, consistency of what you should be looking at.

There should be a discussion around what is in front of you. I think Stuart did a nice job of beginning the conversation around what is in front of you.

If there is general agreement that it meets the quality, quantity, consistency, that would be fine; we could move forward. But, again, in the spirit of having this be the sample of what we want this discussion to be going forward, some additional discussion around whether we think that that has been met and, if not, there are additional options for how you can move forward in terms of general agreement that, even if it is not presented, that you could move forward. But, anyway, that doesn't seem to be the case here. But some discussion seems necessary.

MEMBER PELLETIER-CAMERON: I just have a quick comment, just to reiterate what Nancy said. She was discussing just her concerns about the numerator at the very top. I mean, the numerator, there are two

questions, whether you discussed incontinence with your provider, and there is no mention of who brought it up, which is fine, coming from a urologist perspective.

And the second one, whether you received treatment or not. And I agree with Nancy that you can be offered treatment options, but it is up to the patient to decide whether they want treatment. So, they may have been given their options and opted not to be treated. But, in a general perspective, I don't think that is poor quality, to have not received treatment, but to have discussed it is good care. But you don't have to elect to be treated. So, that is just a general comment about the numerator.

MEMBER MORTON: I have a comment. I like one part of the measure, in that it is pretty open-ended in terms of the treatment options. It is not shoe-holed into surgery or Kegels or, frankly, even weight loss is something that could be used as a treatment

option. So, I like that aspect of the measure, that it is pretty open-ended in terms of treatment.

And it seems pretty straightforward. The question is whether or not patients had some sort of discussion around their incontinence.

CO-CHAIR SAIGAL: Yes, I agree with all of the comments that were made, but evidence is good that -- they talk about the ACOG guidelines and the other guidelines in here that are from SIGN, that there are recommendations of high levels to treat incontinence with these various interventions. So, that seems to be a good link, in my head.

Yes?

MEMBER LIGHTDALE: This is just more of a comment. I guess part of the guinea pig thing here is you have got some people who really know nothing about urinary incontinence, a good thing. So, I am actually learning like literally reading the document.

I guess I would just go on having to really pick through now to sort out, okay, SIGN used these levels of evidence. They talked about 1-plus-plus as high-quality meta-analyses systematic reviews of RCTs. I don't know if that is two RCTs or ten RCTs. Then, I was trying to translate into your metrics.

So, actually, I would say that it is actually important for the developer to go back and really give me five good randomized controlled trials that I could, then, point to. And I don't want it for me necessarily, but if this is in a toolbox for an organization to use, that they would be able to go to those randomized controlled trials. I think that is important for those of us who are new to the field.

DR. PACE: Yes, and I would just say that we really do ask the developers to rely on systematic -- you know, doing a systematic review of the evidence is another whole area of expertise. And so, to ask the

developer to do a primary systematic review is probably not going to be the best systematic review. However, if they want to do a performance measure and there has been no systematic review, that may be where they are starting.

So, that is why we asked them to cite the evidence that they are providing, and in this case they are using two guidelines that did systematic reviews. But we are asking you to apply the criteria based on what they present in 1(c)(8) through the end of the document in terms of the quantity, quality, and consistency.

And the other thing I will just mention about the Evidence Task Force and that rating scale is that it is not just RCTs. There are other types of evidence that can certainly be considered.

One other thing, when you are looking at the evidence, is whether the evidence presented is on point with what the

measure is measuring. And so, there are a lot of treatments that have good evidence behind them. But this, again, is the nature of the patient experience point of view, is that it is just being asked any treatment. So, there is really no kind of whether they are getting the right treatment. That is not part of this particular measure.

MEMBER LIGHTDALE: Sorry, I just feel like I need to clarify. So, there are two guidelines cited here. That is basically the evidence. I don't know either society or how to evaluate whether that is -- and I have been involved in guideline development to the point that I know that some guidelines are well-done and some aren't. Often, it depends on what is the evidence out there.

So, is that enough to cite two guidelines? I am actually looking at 1(c)(8). Or do you want now --

CO-CHAIR BASKIN: I think where we stand on here is that, yes, they cited two

guidelines. In this particular case, these are guidelines that are evidence-based guidelines rather than consensus-based guidelines. I think we have to rely on the folks here or the urology community to tell us whether these are evidence-based guidelines that are the kind of quality that we would expect. I am not, but I suspect they are. I mean, I am not a urologist, but I suspect that these are well-done evidence-based guidelines and have great background.

We are just going to have to ask our colleagues to make that decision for us, and we will live with it, because that is all we have presented to us. That is what we have our expertise for. So, a urologist to comment would be helpful.

CO-CHAIR SAIGAL: And they break out a number of studies in each of these guidelines. So, that may be helpful. And then, in quality and the consistency, they also kind of enumerate different studies as

well that are in there.

MEMBER MARKLAND: I would just like to add that the guidelines presented, one is European. So, it may not always apply here in the U.S. The other one is U.S.-based.

But there are quality measures and studies, randomized controlled trials, that have been done looking at quality measures in primary care settings. And I think that is where we know the treatment and the guidelines for treatment, but to ask these questions and to see in a primary care setting if asking about urinary incontinence helps improve outcomes, I think that is a very important aspect of this.

There has been a randomized controlled trial looking at the improvement in people receiving treatments for urinary incontinence just by asking and discussing. And so, I think there is evidence outside of treatment guidelines to suggest improvements in processes of care.

CO-CHAIR BASKIN: Let's take these last comments and --

MEMBER REYNOLDS: I am actually going to withdraw my comment because it has been answered already.

CO-CHAIR BASKIN: Zahid?

MEMBER BUTT: Yes, I'm sorry we are taking more time, but I think we are sort of giving feedback in the general sense also.

So, I think I sort of come back to that all comment about guidelines. I think Nancy just mentioned that -- you know, I looked at some of the other ones that I reviewed primarily, and there is sort of the common theme that, when a guideline is quoted, and especially if we are not going to treat that as a study and it just says, "Was a systematic review done by the guidelines developer?", it is yes and no. It seems like perhaps there should be another requirement or at least a recommendation for the developer to actually provide the number of studies that

were done as part of that systematic review of the guideline, which is kind of what she was saying.

DR. PACE: Right, and that is what -- and, you know, we will also get some more feedback on the flow of the form -- but the idea is and the instruction is that, if that systematic review is done, to give a summary of the quantity, quality, and consistency in that section, the last section of the form.

MEMBER BUTT: Okay.

DR. PACE: So, you are exactly right, that is what the intention is.

CO-CHAIR SAIGAL: Okay. Johannes?

MEMBER KOCH: So, I think we could spend a lot of time, all day, arguing between high and moderate. And really, what would be helpful, especially as a GI person, is for somebody to say this does not meet a moderate or high. So, if I vote moderate and it is really high, no relevance. But if somebody could give guidance as to, does this not meet

moderate or high, then we can have a relevant discussion. If it is going to meet moderate or high, say it is moderate or high, and then we can after hours discuss over a beer whether it is high or moderate or not.

(Laughter.)

CO-CHAIR SAIGAL: Great point. Okay. With that, in the spirit of that, can we move to a vote then?

Any last questions about what we are going to do in terms of voting?

(No response.)

We are pretty good about that?

MR. WILLIAMSON: We will now vote on the evidence. This is a yes/no.

DR. PACE: No, there are the three options, yes -- and we are not going to quibble over the high, moderate, low at this stage. So, 1 is, yes, it meets the criteria for quantity, quality, consistency. And then, there are two options. No. 2 is, no, it clearly does not meet it, including it is not

even based empirical evidence. And then, the third one is simply, no, because there is not sufficient information for you to even know. And then, we have different options based on those.

CO-CHAIR SAIGAL: Okay. Can we hit the buttons now?

MR. WILLIAMSON: One second. You can begin voting now.

(Vote taken.)

All right. And we have 15 yes; 2, no, evidence does not meet guidance, and zero, no, insufficient information submitted to rate.

CO-CHAIR SAIGAL: Two people voted no?

MR. WILLIAMSON: No. Sorry. Oh, it is zero. Sorry.

MR. AMIN: Okay. The first time around it is going to take a little bit just to get us there.

So, the last one is, the last

subcriteria under importance to measure is -- Evan will get us there -- it is the opportunity for improvement and the performance gap. It is slide 53.

So, what we are looking for here is the data demonstrates considerable variation or less-than-optimal performance across providers and/or population groups. The distribution of performance scores, the number and representativeness of the entities included in the measure performance data, data on disparities, the size of the population at risk. And then, we will use a high, moderate, low, or insufficient rating scale.

And I will turn it back to the group for discussion and then vote.

CO-CHAIR BASKIN: So, comments regarding gaps in performance?

CO-CHAIR SAIGAL: Nancy, any comments about that?

They have a good section in here about the gaps in performance. Where is it?

Section 1(b)(2).

MEMBER PELLETIER-CAMERON: Yes, 1(b)(2) and 1(b)(3).

CO-CHAIR SAIGAL: Page 4.

MEMBER PELLETIER-CAMERON: Yes. There is raw data and there is also several citations that are all pretty unanimous in their impression of the gap that exists.

CO-CHAIR SAIGAL: Yes. And they have health plan data in here as well. So, I would say it is pretty convincing.

Any other comments?

MEMBER MORTON: I would say that it looks like there is a lot of data to support that people don't have this question posed to them very often, for a variety of reasons. It is a sensitive subject. So, it seems like that there is a lot of support that this is not discussed commonly.

CO-CHAIR SAIGAL: Right. Yes. Okay. Good.

Anyone else? You wanted to make a

comment? Oh, no?

Okay. Any other last comments before we vote?

(No response.)

Okay.

MR. WILLIAMSON: We will now vote on the performance gap. There are four options: 1 is high; 2 is moderate; 3 is low, and 4 is insufficient.

You may begin. Once the clock shows -- sorry -- with the mouse, it takes a while to show up. You may begin voting now.

(Vote taken.)

And we have 13 high, 2 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. That's it. That is the process to move to the next one.

MR. AMIN: Well, we will do an overall vote on the whole concept, and then that will be the tenor of the conversation going forward.

CO-CHAIR SAIGAL: Okay. Do we have to hear discussion about this part, too? Or do we just move to --

MR. AMIN: Basically, what you are doing here is that you are going to just recommend approval of this concept, which will then go to the CSAC.

CO-CHAIR SAIGAL: Okay.

MR. AMIN: There is some question.

MEMBER FALLER: Isn't it at this point that we make the recommendation that they should add the data that she was talking about that includes the fact that discussing incontinence with your provider improves --

MR. AMIN: So, this would be a good time to be the checklist, what you would want to see before you would let this measure come forward in stage two. So, if there is additional -- if the measure of submission is missing components that you would want to see before to complete the importance criteria, this would be the time to do that, or if you

have any other considerations before stage two. So, as explicit as you can, just so that that information goes back to the developers.

MEMBER MARKLAND: There is a reference in this that does reference that work. It is not explicitly stated as evidence, but they do reference the work, the earlier work, from this group that looked at quality indicators and vulnerable elderly.

CO-CHAIR SAIGAL: So, you are satisfied with that? Okay. Great.

Do you want to do any summary before you vote or shall I do a summary?

MR. AMIN: If you --

MEMBER GILL: Excuse me. I would also be concerned about that one word where it says, "received treatment" versus "offered treatment". I think that would make it a better measure, given what everybody has said about offering versus -- because the patient can refuse the treatment or not decide on it.

DR. PACE: You might want to ask

the developers to comment on that because this is part of the Health Outcomes Survey. So, I don't know how much flexibility. Is that okay, to ask them to --

CO-CHAIR BASKIN: Any comment at all regarding the wording of the question?

DR. GIOVANNETTI: Yes. That is a concern that we have heard before, and we are working, actually, on modifications to this measure. However, the timing of that work didn't sync up with this Committee.

So, the measure that you are presented with is the measure as it is currently in the Health Outcomes Survey. We hope in about a year, after we finalize all of our testing on the new measure, we will have an improved measure that will get around some of this issue around the received treatment versus did you discuss different treatment options.

However, as we have this measure as it is now, it is in the Health Outcomes

Survey. We can't just go in and change it, based off of the recommendations of this Committee. You know, we can't change it tomorrow.

So, I believe the option in front of you is to look at this measure as it is now. If you still think the concept is worthy of endorsement, we will move forward to stage two. If this is a straw that breaks the camel's back on this measure, then that is your own kind of prerogative. I can only tell you that we are working on improving this measure.

MR. AMIN: So, in terms of a summary, Chris, I think it would be helpful not necessarily for you at this point, because it seems like the general consensus for the group was that it meets all three subcriteria fairly well.

CO-CHAIR SAIGAL: Okay.

MR. AMIN: And I think that has been fairly well-stated.

CO-CHAIR SAIGAL: Time to vote.

MR. WILLIAMSON: We will now vote on the overall recommendation. This has two options, yes and no.

And you may begin voting now.

(Vote taken.)

And we have unanimous approval, 15 yes and zero no.

MR. AMIN: Okay. Evan, if we could switch back to slide 35?

MR. WILLIAMSON: Thirty-five?

MR. AMIN: Yes. Okay.

So, I assume that there will likely not -- and you guys in the group can tell me if we need to not -- we will go through it very quickly now in terms of what you are looking at. But, again, we will start with the importance-to-measure criteria.

Now this will be, again, the nature of how this conversation will go. So, you got a general sense of what -- we will be looking at high impact, then the evidence that

supports the measure focus, and then the gap.

So, the first one is looking at high impact to assess whether the measure focus is evidence-based.

CO-CHAIR SAIGAL: For Measure 0098, and Dr. Baskin here is going to be the lead discussant for that

MR. AMIN: Yes. So, we are looking at high impact.

CO-CHAIR BASKIN: All right. So, Measure 0098, very similar in that we are measuring something very much like the last measure, except that this is an administrative measure of providers and really has to do with documentation in the record and actually submission of claims information that the documentation is in the record, which makes it a little bit different.

But in terms of the importance, it is the same issue of stress incontinence and the information given in terms of the impact is literally identical to the last measure.

So, I am going to just spit it out, that it seems like this is a high impact based on the same, exact item that we just reviewed.

And looking for any comments around the room regarding the impact.

(No response.)

Okay. So, I think that is pretty clear. I think that we could really progress and just move on to voting regarding impact, since it is so straightforward. So, why don't we get the voting machinery up and running?

MR. WILLIAMSON: We will now vote on high impact. There are four options: high, moderate, low, or insufficient.

You will begin voting now.

(Vote taken.)

One person is out.

CO-CHAIR BASKIN: One person hasn't. So, press again, if you haven't.

There we go. Okay.

MR. WILLIAMSON: There we go.

CO-CHAIR BASKIN: And let's see,

and away you go. I don't even get an envelope to open, but there it is. That is the equivalent of an envelope in today's world.

(Laughter.)

MR. WILLIAMSON: We have 15 high, zero moderate, zero low, and zero insufficient.

CO-CHAIR BASKIN: Okay. And the next one is the --

MR. AMIN: So, the next one is that we are going to be looking at evidence.

Evan, if you could go to slide 46?

And so, basically, you are going to look at the quantity, quality, and consistency of the evidence that is presented in front of you. Ideally, if you can go through each of those --

CO-CHAIR BASKIN: Which one was first? Help me out there.

MR. AMIN: Oh, the quantity.

CO-CHAIR BASKIN: Quantity? Okay. So, just to lead this discussion, once again,

this so parallels the prior one. I mean, the evidence is the same evidence, okay, in terms of the quantity of evidence. And once again, it is those two large meta-analyses that we accepted last time around.

So, is there any concern about those, the body of evidence in terms of quantity for this particular measure that we would want to speak about? Anybody want to bring that up?

Go ahead.

MEMBER BORDEIANOU: We are looking at the evidence that urinary incontinence needs to be treated. And there is a lot of evidence on that. But does documenting your medical record, when you have a 15-minute visit as a PCP and you have discussed hypertension, and so on and so forth, really impact how well a patient is treated for incontinence? And should we be measuring the PCPs on that?

CO-CHAIR BASKIN: Yes, so the

concern here is kind of similar to the last one in that there is evidence that the treatments work. There is a large body of evidence to that. There is evidence that there is under discussion of this particular problem. But the question is whether improving on this measure actually improves outcomes. That is the leap of faith that we are making here, and the leap of faith we made on the last one, too.

I don't know. That was my comment in the pre-discussion, was, you know, where is the evidence linking the measure itself to a better outcome? And I don't know that that exists. But if someone has a comment on that, then go ahead.

Chris?

CO-CHAIR SAIGAL: I could ask our colleague -- I don't have your name of me. I'm sorry, what was your name again?

MEMBER MARKLAND: Alayne.

CO-CHAIR SAIGAL: Alayne, you had

some data about a rural implementation of this?

MEMBER MARKLAND: Well, it is broader quality care for older adult data where urinary incontinence is one of the measures of improving care for all older vulnerable adults.

And in that, though, they do use some administrative data, which this is very different from the last measure, that it is all based on administrative data, which we know tends to underestimate rates of discussion. Especially, I am questioning some of these CPT codes to actually measure the assessment of versus an ICD-9 code, which would be the only character here.

I think it is still important to assess how often these are being coded in a primary care setting to determine, does it, therefore, link to treatment. So, I think there are two, the patient and the provider level. This gets more at that provider level,

in my opinion, which is a part of the other study I mentioned.

CO-CHAIR SAIGAL: In the other study you mentioned, the assessment in the primary care setting was linked to treatment, is that right?

MEMBER MARKLAND: Uh-hum.

CO-CHAIR SAIGAL: So, there is some data about that, then?

MEMBER MARKLAND: Uh-hum.

CO-CHAIR BASKIN: Other comments?

(No response.)

Then, we should move on to a vote regarding the -- wait. Do we do the quantity of evidence?

MR. AMIN: You do the overall -- I mean, so you are going to discuss quantity, quality, and consistency.

CO-CHAIR BASKIN: Oh, all at one time?

MR. AMIN: And then, you vote all at one time on the body of the evidence,

whether it meets all three of those.

CO-CHAIR BASKIN: And then, any other comments? Because, I mean, that is the one glaring issue here, is the relationship of the measure itself to improving outcomes. I think we were pretty clear last time that the evidence shows that there is an issue, and it is not discussed, and that the treatments are valuable. So, that is the one leap of faith I think we are making based on some evidence that wasn't submitted.

And any reason why we shouldn't go to a vote?

(No response.)

I guess not. Let's go to a vote then.

MR. WILLIAMSON: We will now vote on the evidence. There are three options: yes; no, the evidence does not meet the guidance for quality, quantity, consistency, or, no, insufficient information was submitted.

You may begin voting now.

(Vote taken.)

And we have 13 yes; zero, no, that the evidence does not meet the guidance, and 2, that insufficient information was submitted to rate.

CO-CHAIR BASKIN: Just so that we are all clear, anybody want to make a comment other than what was already suggested as to why the evidence may be insufficient? Is there some other reason it was thought to be insufficient, just so we can document that? Or was it simply the link between the measure and the outcomes, which is, I think, something we have already discussed wasn't there?

(No response.)

Okay. No one has to make comment. I just wanted to make sure there was nothing that was missed in the conversation that we should have documented.

MR. AMIN: And now we go to performance gap.

CO-CHAIR BASKIN: So, performance gap, they did give us information here in 1(b)-(c), and this is from the PQRS, of which this measure is a part of PQRS.

The issue here, I think, is rather obvious, in that the percentage of providers, eligible providers, who actually report is so low that the performance gap is based on a very, very small sampling of providers. And the performance gap seems to vary from year to year rather significantly, but probably because of the low volume of providers, there is probably a lot of randomness here.

I think, though, that some of the evidence review and some of that evidence does actually speak to a performance gap other than this PQRS information. So, my personal feeling is there is a significant performance gap here based on that evidence, the body of evidence that was submitted, as opposed to just the PQRS data. I think the PQRS is just rather insufficient to make any conclusion on,

but that is my personal opinion/review. But I am opening the floor to any other comments regarding the performance gap.

MEMBER BUTT: So, the PQRS data, could you interpret by just saying that there is not enough overall compliance?

CO-CHAIR BASKIN: Well, the way I see it is the compliance rates, you can see they vary anywhere from the high 50s to 90-some percent, and from year-to-year it changes considerably. But, if you see, it is based on such a small number of physicians --

MEMBER BUTT: I mean, is that the interpretation, that there is overall less compliance or overall less reporting?

CO-CHAIR BASKIN: Well, I think it is both. So, if you felt that this was a reasonable sample of physicians, then you would say the compliance rates vary in the high 50s to 80s. I think that is a reasonable-enough performance gap, and I suspect it is much higher than that because

this is a selected population who voluntarily submitted this information. They didn't have to report on this particular measure. I am pretty certain that folks who don't do so well probably didn't voluntarily report.

So, I personally don't think the PQRS data is very helpful for me in terms of the evidence of a performance gap. To me, it is all in the literature that backed up the evidence reviews, is really where the performance gap is demonstrated.

MR. WILLIAMSON: We will now vote on the performance gap. There are four options: high, moderate, low, or insufficient.

And you may begin voting when the mouse shows up. There it is.

(Vote taken.)

CO-CHAIR BASKIN: I feel like we are on Jeopardy here, and everyone is trying to be the first to click.

(Laughter.)

MR. WILLIAMSON: And we have 7 high, 8 moderate, zero low, and zero insufficient.

CO-CHAIR BASKIN: Okay. So, to summarize that, I think it was pretty clear, the importance of the issue. That has been well-demonstrated. And I think we have clearly showed that there is some evidence, strong evidence, that there is adequate treatment and that there is certainly adequate evidence submitted.

The question only became the evidence regarding the actual measurement process itself and how that actually relates to the outcomes, as to whether actually the measure will improve outcomes in and of itself. I think that gap exists, and it is may exist in the literature. There seems to be only a little bit of information out there.

The issue with the -- what is the last thing we just voted on?

CO-CHAIR SAIGAL: The performance

gap.

CO-CHAIR BASKIN: Oh, the performance gap, understandably, many of us thought moderate versus high was probably based on the idea that PQRS data is rather insufficient, and I think that was pretty clear to us. But, overall, I think that moderate, high, everyone considered that, that there is enough of a performance gap here to support this particular measure.

So, I think we can go ahead and take a vote on this measure.

MR. WILLIAMSON: We will now vote on the overall recommendation. There are two options, yes and no.

And you may begin voting now.

(Vote taken.)

And we have 14 yes and 1 no.

MEMBER BUTT: Could I make a suggestion at this point for the developers, like we did before for the previous one?

CO-CHAIR BASKIN: Yes. Yes,

please do. In fact, I was going to ask that.

MEMBER BUTT: Oh, I'm sorry.

CO-CHAIR BASKIN: But, even before that, I was going to ask if anybody wanted to -- and this is purely voluntary, once again -- but if anybody wanted to give the dissenting opinion, we would be welcome to do that for the documentation, but there is no obligation to do that.

MEMBER BUTT: I was just going to say that we would like to encourage them to look at an e-measure specification for this measure because it looks for the plan of care in the numerator, which is currently defined through CPT II, which is sort of an administrative, by definition, claims type of plan-of-care definition.

So, I don't know if they have plans to try to specify that more, so that EHR data could be used to get at the numerator.

CO-CHAIR BASKIN: Well, I do think, in general, almost all measures at some

time will go through the e-measures process over time. I think all new measures, don't they have to go as well?

DR. PACE: Well, I will let Heidi speak.

MS. BOSSLEY: We have not yet instituted a requirement that -- most measures, not all measures, can, of course, be translated into EHR data. But we don't have that requirement yet, in part, because the tools that the developers would need to get it into the standardized format and a few other things is probably not quite ready for them.

So, the hope is that over time we would see measures move forward, and exactly this type of measure you see in front of you would be perfect for translating into an e-measure. And so, that would be the hope. But there is no requirement yet.

CO-CHAIR BASKIN: I have one comment for the developers. It is similar to, basically, the other one, but I don't think

the hurdle is quite as high on this one as it is on the survey and changing it, in that there really should be an option for member choice of no treatment, if there isn't already. I don't believe there is.

I don't think that would be -- it is unclear to me on the plan of care that the plan of care includes the conclusion "discussed with member; member chose no treatment as the plan of care," but that should be an option available. So, just a suggestion to take back to consider, please.

Any other comments on this before we close out on this measure?

(No response.)

Okay. Well, we went a little over to get this done from our break, but we will still get a break. So, how long of a break will we take? Still take 15 minutes or not? All right, we will take a 10-minute break. About 11:15, we will resume.

Thank you.

(Whereupon, the foregoing matter went off the record at 11:04 a.m. and went back on the record at 11:17 a.m.)

CO-CHAIR BASKIN: Folks, we are convening again.

Just one brief comment. We did it a little bit on this first time around for these first two measures, but I understand that everyone has concerns, when they read these measures, about the construct of the measure, the ability to measure; is it really measuring what we think it is going to measure and really have the same outcome?

So, we need to understand that in this process the validity/reliability of the measure, the ability to measure what it is supposed to measure really comes up in stage two. But it is not unreasonable at some point to make a suggestion or a comment after the voting is completed.

So, in other words, we complete the voting on the part we are talking about,

the importance. But, then, after that, I mean, if anybody has a comment that they would like to send back with the developer or into the record of concerns that, gee whiz, if you don't really think about this between now and stage two, you may have trouble with validity and reliability, I think some comments are reasonable.

What I don't want are those comments to be part of our discussion and vote because we don't want to get into those weeds in this stage one of the process. So, for instance, when we made a comment about, you know, a change in the Healthy Outcomes Survey, or a potential change in that, as a suggestion, that is a reasonable thing to do after we are done with the voting. Let's limit that, though, not a lot of time spent on it. But, as we get into some of the newer measures, those suggestions are insights may be helpful to the developers. So, when they come back, they will actually have a valid and

reliable measure.

CO-CHAIR SAIGAL: Great. So, maybe we can start off with the next measure set from the AUA. Can the AUA presenter please give us two or three minutes of background?

MS. WILBON: Excuse me. I'm sorry, if you could just give an overview of all of your measures, instead of a one-by-one approach, that would be helpful for all.

MR. CLEMENS: I am happy to do all five. That was my plan.

MS. WILBON: All five, actually, yes. Go ahead.

MR. CLEMENS: So, I am Quentin Clemens. I am a urologist at the University of Michigan. I am here on behalf of the AUA to discuss the five measures that we have brought to your attention today.

This measure set was developed under the auspices of the AMA Physician Consortium for Performance Improvement Work

Group which was jointly chaired by the AUA as well as the American College of Obstetrics and Gynecology. It included members from family medicine, geriatrics, and nursing.

This is a measure set that we have planned to submit to the PCPI shortly for their comment and potential revision. With this process here at the NQF coming up, we put that on hold to bring it here first.

The focus is on the surgical treatment of stress incontinence. We know that there are between 100 and 200 thousand of these surgeries done every year in the U.S. So, it is an important thing and a major component of the practice of urologists and urogynecologists.

We have five measures. The first has to do with a complete evaluation of patients. The denominator for all of these is women who had stress incontinence surgery.

And the important thing with measure 1 is certain components, such as a

physical exam, assessing a post-void residual. We know that these surgeries often make it harder for the bladder to empty afterwards, and it is important to assess that, as well as objectively showing that the patient has stress incontinence.

Anecdotally, at least many of us see patients who have failed surgery and it is pretty clear they didn't have stress incontinence all along. So, to be able to really show that we felt was important. These are all based on evidence-based guidelines from the AUA, I should mention.

Then, we have two counseling measures. One is making sure that the patients who have had surgery were counseled about other options, such as behavior therapy and medical therapies.

The other counseling measure has to do with patients who have had mesh placed for stress incontinence, to make sure that they were counseled about that. And in

particular, there are three components to the counseling: letting them know that erosion can occur, that pain can occur, and the mesh is permanent.

Measure 4, then, has to do with performing cystoscopy at the time that the surgery is done. We know that these mesh slings, or any type of sling, actually, can be placed into the bladder. If that is recognized interoperably, it can just be repositioned without any problem to the patient. But if the cystoscopy is not performed and there is some foreign material in the bladder, then that can cause major morbidity for the patient postoperatively. So, using cystoscopy during the surgery is the fourth measure.

And finally, we have a followup measure, measure 5, indicating that patients who were characterized postoperatively with many of the things we recommended preoperatively, such as a pelvic exam, post-

void residual analysis, and, of course, assessment of their symptoms.

This was indicated to be within 12 months followup. It is a reflected, a significant discussion among the Work Group about differences in practice patterns, where some may at the short-term simply do a telephone call and then ask the patients to come back at a longer followup. And others in the group, based on practice patterns, would have patients come in earlier in person and then have them call later if there were a problem. So, that was the reason for the 12-month time period for measure 5.

And with that, I will turn it back to you.

CO-CHAIR SAIGAL: Thanks, Quentin.

Okay. Paul, could you lead us through Measure 2049, starting with importance to measure and report, high impact?

MEMBER MERGUERIAN: Measure 2049, basically, deals with the preoperative

assessment of patients with urinary incontinence. Basically, their denominator includes history and physical, testing performed, and, also, evaluation performed.

The AUA has come up with some guidelines which I have reviewed. And you can actually look at the AUA site. There are basically four chapters there. One of the chapters deals with evaluations of patients with urinary incontinence, with references.

CO-CHAIR SAIGAL: So, in terms of the impact of the measure that we are talking about, which is complete workup of stress incontinence, this is a big problem, you would say, that people aren't doing this, based on the evidence that they have presented?

MEMBER MERGUERIAN: So, the evidence that they presented was that it is a significant problem. Millions of women are affected. The impact of SIU treatment in the United States exceeds $13 billion per year, and around 30 percent of women describe their

measures are being bothersome.

CO-CHAIR SAIGAL: Okay. Any other comments?

MEMBER MERGUERIAN: It is high impact.

CO-CHAIR SAIGAL: Incontinence is a high-impact problem?

MEMBER MERGUERIAN: The incontinence is the high-impact problem. The fact that the management, I mean, there is really not a lot talking about individuals paying themselves for care costs related to urinary incontinence, and that is a huge impact on the cost to the individual, but also on the cost to the healthcare system.

CO-CHAIR SAIGAL: And there is data in there about the lack of performance of the complete physical exam and the --

MEMBER MERGUERIAN: There is data in there that suggests -- I mean, again, the numbers are pretty small -- but there does suggest that there is quite a gap, where

around 66 percent of those surveyed -- and this is not just from the urological literature, but also from gynecological literature. I mean, the gap is around 66 -- there is compliance with around 60 to 80 percent.

MR. AMIN: Chris, could I just make one quick suggestion?

CO-CHAIR SAIGAL: Yes.

MR. AMIN: Again, ideally, we would just do high impact, and then the evidence, and then the performance gap. It just keeps the conversation a little cleaner.

CO-CHAIR SAIGAL: Okay.

MR. AMIN: So, if everybody is okay with high impact, just feel free to move on, if that is where everybody feels, but try to keep the conversation that way, so that it keeps it a little more structured.

CO-CHAIR SAIGAL: Okay. So, then, any other comments about the impact concern?

(No response.)

Okay. So, shall we vote?

MR. WILLIAMSON: We will now vote on high impact. There are four options: high, moderate, low, or insufficient.

And you can begin now.

(Vote taken.)

And we have 13 high, 2 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. And now, we will talk about the quality of evidence.

MEMBER MERGUERIAN: I guess the AUA guidelines, basically, they did the meta-analysis. They reviewed over 7,000 articles and came up with 150 articles that they based their guidelines on. So, there is quite a bit of evidence relating to the evaluation of these patients initially.

I think the eventual impact, it could be also a reduction in healthcare cost where, overall, in the long-term you may reduce costs of surgical treatment if the incontinence is actually characterized prior

to surgery.

CO-CHAIR SAIGAL: Any other questions or concerns about the evidence?

(No response.)

MR. AMIN: So, again, I just would like to have the Committee keep in mind that you should rate the information that is presented in front of you in terms of the quality, quantity, and consistency, which is provided in 1(c)(8), 1(c)(9), 1(c)(10), and 1(c)(11). And maybe I will leave it there.

MEMBER MORTON: It seems like this is kind of an appropriateness question more than anything else and making sure you are operating on the right patient, the right indications.

From what I could see in the data, it was just surveys, like how often people were doing it. I don't know if there is more enlightenment around appropriateness, if there have been any studies around that.

CO-CHAIR SAIGAL: Well, we have

two incontinence specialists I am aware of at least in the room from urology. I don't know. Do you guys have any comments about other data about this topic?

MEMBER PELLETIER-CAMERON: I am aware there is more data that is not even referenced here about surgical -- it is a randomized controlled trial that was published on the results of two different incontinence surgeries in women. The group that had the worst outcomes in terms of care were the people who they never actually demonstrated their incontinence. So, if you couldn't prove they had incontinence, those people did badly. No kidding, yes.

MEMBER MERGUERIAN: And the other thing, if you look at the guidelines and actually go to their website, they actually have a table with all the different treatments that are provided and, then, also, the success rate of these with CIs.

CO-CHAIR SAIGAL: So, it sounds

like specifically about the idea of whether doing these interventions improves the outcome, there is some evidence that may not be cited in this document, the summary? Okay.

Any other comments before we vote then?

MEMBER BUTT: So, are we voting for the evidence presented and then --

CO-CHAIR SAIGAL: Yes, there are three options. Either you vote that the evidence meets the guidance there or it does not, or it is insufficient -- there is data, but it is not submitted, basically.

MEMBER MERGUERIAN: That is for the first concept, which is a complete workup.

CO-CHAIR SAIGAL: Right, for 2049.

DR. PACE: Right. So, this particular measure is doing the workup. I know the comment was made about 7,000 studies. I doubt that the 7,000 studies were about doing the workup. That is what we have asked for each guideline. You know, this is based

on a guideline recommendation. I think it is primarily -- you know, whether that is evidence-based or expert opinion, but we ask for a summary of the quantity, quality, and consistency of the body of evidence for that particular guideline, not in general for a whole clinical practice guideline document that may have multiple recommendations. So, the evidence for this particular recommendation.

CO-CHAIR SAIGAL: Go ahead, Stu.

MEMBER REYNOLDS: Well, to echo, though, specifically, if you look at the evidence that is submitted, there, admittedly, is probably not that great evidence that shows, specifically looking at this, whether you had an assessment that changes the outcome. There is certainly strong consistency across these guidelines and, then, other ones about the appropriateness of what needs to be done before you do surgery.

Certainly, anecdotally, from a

personal experience, I see a lot of patients who come in who have complications. And you ask them, and they have never had a physical exam before they had surgery, which is a bit egregious, but there is no data to support that, and that is just anecdotal.

I guess I would support the consistency of it. But, specifically, the AUA guideline is the standard, which is based primarily on consensus as opposed to strong data. I don't know that -- they don't clearly demonstrate how many, in terms of quantity of the studies that are here, but I think the consistency certainly is highly strong across all the guidelines and the references that they give.

CO-CHAIR SAIGAL: So, to summarize what you said, basically, in terms of consistency, what data are out there suggest that it is an important thing to do, but there are not many studies that are out there. The guidelines are based on consensus to a large

degree. And there is at least one study that we are aware of that is not submitted here that does point in the same direction. Is that fair?

MEMBER PELLETIER-CAMERON: There is never going to be a randomized controlled trial about whether working up your patient before you do surgery or not impacts their outcome because of the ethics involved of doing surgery on people without doing a proper workup. So, I don't think the data would ever become available that answers that question.

MEMBER MERGUERIAN: Yes, and they state that in their link, that it is basically based on consensus statements. Those recommendations for evaluation are based on consensus.

CO-CHAIR SAIGAL: Yes.

MR. AMIN: So, Chris, can I provide some process guidance here?

CO-CHAIR SAIGAL: Sure.

MR. AMIN: So, what you would want

to do in this case, I mean, really, my understanding of the summary statement of the Committee is that the information presented here would not meet the NQF criteria for quality, quantity, and consistency. But once you vote to, if you vote to, I should say, you will have the option of saying that there is information out there that would meet this criteria, if I am hearing the Committee correctly.

CO-CHAIR SAIGAL: Is the one that you would use for that purpose?

So, to be clear, if you say three, that there is in the document not enough evidence, but there is evidence that exists that could be put into the document, if you said three? Or you could say one, and, basically, you feel that consensus statements are enough. That would be an exception, though, right?

DR. PACE: Right. Consensus statements are not what we need to consider

meeting our evidence criterion. So, that would be No. 2. It doesn't meet our criteria, but you could consider it as an exception to the evidence criterion.

CO-CHAIR SAIGAL: For example, if we got a vote where everyone said two, we could then see if there was room for an exception because of the special nature of these kinds of measures. Okay.

So, with that, are we ready to vote? Okay.

Question?

MEMBER SCHOENFELD: So, I may have missed this, but I was going to ask, could the lead discussant -- or is this not the way you want to do the process? -- actually give their recommendation about how they would vote on this particular issue?

CO-CHAIR SAIGAL: I think that is a great idea.

MEMBER MERGUERIAN: I think I would vote for two, that it is basically a

consensus and there is no evidence that it does not meet the guidelines or the guidance for quantity, quality, and consistency.

CO-CHAIR SAIGAL: Okay. And as I said, I would probably vote three because there is some data out there that we could introduce, it sounded like to me. So, that is what I would say.

Stu?

MEMBER REYNOLDS: True, but, then, I guess my question is this option that we have to throw out the exception and then push it forward even with insufficient data, but, clearly, the benefit outweighs the harms and all that kind of stuff. Where does that come out? And can we have a preliminary thought that, if we all voted three, that it was insufficient, but we thought it was important enough we would also have a vote on that? Or, from a procedural standpoint, I guess, where do we look at that?

DR. PACE: Yes, yes.

CO-CHAIR SAIGAL: Go ahead.

DR. PACE: Yes, if you vote it down on either two or three, you can bring up -- you know, if someone wants to say, yes, but evidence does exist, we have a way for you to vote on that to discuss it and vote it, or if you say, even though it is consensus opinion and not evidence, we think it is important, then there is a way for you to vote on that to move it forward.

So, the whole point of this vote is one of the key principles of the Evidence Task Force was to be very transparent about the evidence that does or does not exist. And then, you, as the Steering Committee and the experts, can either say, you know, it calls for an exception to having empirical evidence or there is additional evidence that exists that wasn't represented.

But we just want it to be very clear to all the constituents who will then be also reviewing your work and your

recommendations to understand on what basis you moved it forward.

CO-CHAIR SAIGAL: Thanks.

Okay. So, with that, let's vote.

MR. WILLIAMSON: We will now vote on the evidence. You have three options. The first is yes; the second is, no, that the evidence does not meet the guidance for quality, quantity, consistency, and three is, no, that insufficient information was submitted to rate.

You may begin voting now.

(Vote taken.)

And we have zero yes; 4 for, no, that the evidence does not meet the guidance, and 3, no -- sorry -- 11 for, no, insufficient information was submitted to rate.

CO-CHAIR SAIGAL: So, process-wise, then, what do we do now?

MR. AMIN: Okay. So, because now you are at three, you will describe again, just for the record, so that we are

understanding, the evidence that does exist. And then, you will vote on whether there is general agreement within the group that the quantity, quality, and consistency of the evidence, based on the evidence that is discussed that does exist out there, that it would meet the quality, quantity, and consistency of the NQF criteria.

CO-CHAIR SAIGAL: Okay. So, we heard from Anne that there was at least one study that showed that women who had an incomplete workup as defined by the measure had worse outcomes?

MEMBER PELLETIER-CAMERON: So, not even an incomplete workup. This is a subset analysis of patients where they couldn't actually demonstrate their incontinence, fared poorly after surgery.

So, it is one more piece of work, but it is not a randomized controlled trial concerning this.

CO-CHAIR SAIGAL: Okay.

MEMBER PELLETIER-CAMERON: I think I more mentioned it just for completeness' sake.

CO-CHAIR SAIGAL: Okay. So, then, it sounds to me like maybe we will have to consider this as an exception then. I mean, the issue in front of us is whether there will ever be data developed to look at whether these interventions before surgery are going to be impactful in terms of their outcome. And probably my feeling is that no one is going to fund that kind of trial. As Anne mentioned, it is probably not ethical.

So, we have to decide if it is important enough from a population health point of view that these interventions get done, which are based on what experts in the surgical field tend to think that should get done to override the lack of evidence supporting their opinion.

Any comments about that?

CO-CHAIR BASKIN: So, in my

opinion, it is not only that it will never be done, but it doesn't even need to be done, frankly. I think at some point as physicians we can be honest and sit there and say, hey, you know, an appropriate assessment prior to surgery is kind of like mom and apple pie here. It is one of the basic tenets of medicine in itself. So, to me, I don't think that assessment should ever -- I mean, I don't think that particular study should ever be done. I don't think it is really a question that we have to have. I think we can accept this as basic scientific way of performing medicine.

MEMBER MERGUERIAN: I agree. I think it is the standard of care. It is pretty much what you do when you evaluate a patient that comes into your urology clinic. I mean, to get an analysis, you take a history.

MEMBER MORTON: I agree. I mean, this is very similar to like doing these

applications without having a pH probe ahead of time. You have to have the appropriate indications.

And I agree with Andy that no trial is necessary for this. We didn't have randomized trials for parachutes, either. So, we figured that would work. So, this makes pretty good sense.

CO-CHAIR SAIGAL: Okay. So, then, Karen?

MEMBER PELLETIER-CAMERON: And from a clinician's point of view, what they are talking about doing as a complete evaluation is not rocket science. They are talking about a history, a physical exam, just demonstrating that they have incontinence. We are not talking about complicated, expensive testing. We are talking about a simple office evaluation and making sure they don't have a urine infection, and that is about it. So, it is pretty straightforward.

DR. PACE: So, one of the things

-- and I know this is a relatively-new area for you or for us to do performance measures -- but one of the things that also comes up in terms of where NQF is moving with performance measures and our guidance is that this kind of standard of care, does it rise to the level of needing a national performance measure?

So, we have had this shift in our overall portfolio. It is that assessment measures that are at the very beginning of a long line of steps that had to happen to have an effective outcome is why not measure things that are closer, more proximal to the desired outcomes.

And you may say that this is such a new area or people are doing so poorly with just doing basic assessments that you all consider standard of care, that we need a performance measure. But some of the other things that I have heard you say was about appropriateness.

So, for example, would a measure

of the patients who actually received surgery, and did they have the appropriate indications -- and I don't know if that is even something to consider at this point. But it is something, as you look at these measures, remember, we are talking about a national performance measure, and the things that are most likely to improve patient outcomes and improve the quality of care. So, you need to think about that, but also in the context of where this field is in terms of where performance measures are needed.

CO-CHAIR SAIGAL: So, in that light, Paul, can you make a comment on the performance gap in terms of whether that -- what we are talking about here, basically, is like whether it is a big enough problem. It gets back to that.

DR. PACE: Actually, probably that is the next criterion. And the question here is whether this is an exceptional circumstance that you want to go on and evaluate that

performance gap. Because if it doesn't meet the performance gap, it won't meet the criteria anyway.

CO-CHAIR SAIGAL: Right.

MEMBER SCHOENFELD: Just a quick comment? Because it almost sounded to me like it is almost premature to answer this question until we hear about the performance gap because wouldn't a significant performance gap be the reason to answer yes to this question? Maybe I am missing that.

DR. PACE: That is a good point.

CO-CHAIR SAIGAL: Well, I think we talked about it to start with. But, basically, what you were saying was, is this an extreme enough problem? Like maybe we could revisit how big of a problem it is because we are basically saying we are not going to need any evidence that you would qualify as evidence to measure it. I mean, maybe we could reconsider the magnitude of it.

DR. PACE: So, I think this is

question is, do you want to move on, continue to keep considering this, even though it is a consensus-based guideline recommendation and it is kind of the first step in a long line of things that happen? But this is just kind of a preliminary, yes, we think it is important and we want to move on to the next criterion.

But I will defer to you. If you want to have that discussion about the opportunity for improvement, I think that is perfectly fine as well.

CO-CHAIR SAIGAL: Okay. So, does anyone have any comments about the -- go ahead.

MEMBER BUTT: I think it will get trapped in the next one, if it makes it through here. So, it would be fine to go ahead.

CO-CHAIR SAIGAL: Agreed, but the whole point is to avoid having them go down a path of developing something that we don't think is important.

MEMBER BUTT: No, I mean, the gap, it will be caught in 1(b).

CO-CHAIR SAIGAL: Oh, I see what you are saying. Okay.

Okay. So, then, I guess maybe we could just move to the exception vote then. Okay. Is there a special thing we can put up there for that? All right. So, let's vote on whether this meets the qualifications to be important enough to be an exception to the evidence rules.

MR. WILLIAMSON: We will now vote on the exception. And the question is, is there an exceptional and compelling reason that the measure should be considered further? We have two options, yes and no.

And you may begin voting now.

Shall we do a hand vote? And then, I will look at this in the break. It was working before, but it is not working now. We will get that fixed, but we will just do a hand vote.

MS. BOSSLEY: So, this will be good practice for tomorrow because you lose the options of doing electronic voting. The other Committee will be using it.

So, we are going to have you do a hand vote. I'm sorry.

(Laughter.)

Just when you got really good at it, yes.

MR. AMIN: All those voting yes?

(Show of hands.)

Those no?

(No response.)

DR. PACE: So, the next one is ht performance gap.

MR. AMIN: Let me just say that the vote on that was 15-to-0, just so we have it in the record.

Okay. So, performance gap, you are looking at the data demonstrating considerable variation or overall less-than-optimal performance.

CO-CHAIR SAIGAL: Okay. Paul, I think you are going to take us through performance gap next.

MR. AMIN: Paul, can you use your microphone, please.

MEMBER MERGUERIAN: There are several articles that they cite. They are basically articles from the UK, the U.S., Australia, New Zealand, and Canada, that show that there is variability in evaluating patients with stress urinary incontinence. There are some studies from the UK that show that there was a compliance of 10 percent, that they actually performed continence surgery without proper evaluation. There are some other areas where it is as high as 72, in some places where it is as high as 80 percent. But there is quite a bit of variability in the articles that they presented as far as evaluating those patients.

There are 2006 NICE guidelines that pretty much talked about, looked at

responders, and there is really no agreement among responders that some of these tests are actually necessary.

CO-CHAIR SAIGAL: So, how would you characterize this for us in terms of performance gap?

MEMBER MERGUERIAN: That there is a performance gap. There is quite a bit of variation in the way people evaluate these patients.

CO-CHAIR SAIGAL: And that the European studies are probably applicable for the U.S. setting, you would think, in terms of how that --

MEMBER MERGUERIAN: I would think so. They cited the Dutch study that pretty much looked at primary care providers, but most of the other studies were urologists and gynecologists.

CO-CHAIR SAIGAL: Any comments from the Committee?

Liliana, go ahead.

MEMBER BORDEIANOU: Just reading through this, there are a couple of things that are quoted as performance gaps. One, did the patient answer a questionnaire? Two, did the patient get urodynamics testing?

But what we don't know is which one of these specific measures really matter. Is it necessary to do urodynamic testing or is it enough to determine leakage on a physical exam?

So, I am not convinced that the papers that are quoted are addressing these five things need to be documented in each exam, going back to the fact that there is no data because some of this has become standard of care without accumulating.

So, is the role of NQF to say this is standard of care and we want this done, because it is? Or do we ask the societies to go back and dig up material that they will never be able to get anymore because there is some variability?

CO-CHAIR SAIGAL: It is a good summary of the problem, I think.

Zahid?

MEMBER BUTT: And this may, again, be one of the broader questions as well as pertaining to this measure in terms of when a gap is sort of trying to get at the practice of medicine. How relevant do international studies become in that context, when you are trying to do a gap analysis?

Because one can sort of relate to those in a more sort of non-practice type of setting where it is patient sort of symptomatology, et cetera. But I think in terms of where it is trying to get at a procedure or a practice, in that context, it seems like at least a couple of the studies that I can see appear to be U.S.-based studies, and the gap seems to be more not in sort of general H&P, but those urological studies that are specific to measuring the stress incontinence.

So, I don't know whether this is the --

CO-CHAIR SAIGAL: Yes, thank you for your comments.

MEMBER BUTT: -- level of what we are looking for, but at least there are two of them that point to fairly low rates of those assessments.

CO-CHAIR SAIGAL: So, I could summarize, basically, Paul feels that there is enough evidence from international studies to indicate there would be a problem in the U.S. And I think Liliana made the point that a lot of the studies specifically look at elements that are not part of the measure in terms of performance gap, but that she recognizes that they may never have those kind of data. And Zahid mentions that the transportability of those observations into the U.S. setting not be there.

MEMBER BUTT: There may be a couple of U.S. studies that do point towards

a gap.

CO-CHAIR SAIGAL: Yes. Okay.

MEMBER MERGUERIAN: I think one of the other things that was not mentioned -- and maybe it should be mentioned -- in this, which is when you actually look at the AUA guidelines, they give examples of things, for example, if you have a patient that has high post-void residual and has stress urinary incontinence, you might decide a different type of treatment for that patient, a different type of surgery.

And so, those are things that basically maybe should be included in that concept, also. That may guide the urologist to manage this patient differently.

CO-CHAIR SAIGAL: Okay. All right. I mean, I think we have some other comments about this. We should probably vote about the demonstration of the performance gap.

Ready to vote?

MR. WILLIAMSON: You will now vote on the performance gap. There are four options: high, moderate, low, or insufficient.

And you may begin voting now.

(Vote taken.)

And we have zero high, 13 moderate, 2 low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. The final issue is really approval of the concept. And just to summarize, we have decided or the group has talked about that this is an important concept to measure. The evidence supporting the specific measure set was insufficient to meet NQF standards, but we felt that the overall problem was a big enough deal that we would go ahead and make an exception. In terms of the definition of the gap, there is moderate evidence that the gap existed. And now, we are going to see if you want to approve the concept overall for the measure to proceed, with the exception that we

noted.

Okay? Any comments?

(No response.)

Okay. Let's vote.

MR. WILLIAMSON: You will now vote on --

CO-CHAIR BASKIN: It is more of a question because I think I am understanding what some folks here who are urologists -- and I am not; as an internist, I am not sure I have the answer here.

So, I guess what I am trying to understand is, yes, I see there is a gap, that these things aren't being done. What I am trying to get is the link here. Does this evidence include the information that, had you done this preoperative assessment, you would have had a better diagnosis and performed the better or the right type of surgery? I am having trouble with that link.

If you did this urodynamic testing, does that lead to a better outcome,

a better surgery, the right kind of surgery, or something? Or are we asking people to do an assessment that I have no proof is of value?

MEMBER PELLETIER-CAMERON: There is data to show that, if you were doing stress incontinence surgery on people who have, for example, urge incontinence, which is not stress incontinence, they do very badly because they had the wrong kind of incontinence.

And stress incontinence surgery is, I mean, as a general rule, well-paying and easy to do, and urge incontinence is more difficult to treat. So, I mean, in clinical practice I see people all the time who have clear urge incontinence who were treated surgically inappropriately.

And there is data to support that, if you really have stress incontinence, that surgery will work. But there is a big difference between the two types of

incontinence, and determining the difference between the two does involve a workup, does involve a history, a physical, and some testing. You can't just guess.

I think there is a big gap that may not be reported in the literature, but it is present in clinical practice, that people are putting slings in people who shouldn't have them done.

CO-CHAIR SAIGAL: So, I think that, basically, the issue Anne mentioned before is that they are not -- this measure is not recommending a lot of testing be done. I mean, there is a lot of question about urodynamics and other elements of care. It is a pretty basic toolkit that is being required. It is a judgment call about how basic "basic" is, but I think it is pretty basic in terms of our field.

MEMBER REYNOLDS: I would echo that. I mean, you specifically said urodynamics, which is not what is at issue

here. That is a separate level of intensity and workup. And there is some recent randomized controlled trial looking at that, but that is not what is included in here, exactly as Chris said.

So, this is very basic stuff. If any of you have not been to the urologist's office, you get all these when you walk in the door almost every time, in theory, but maybe not in practice.

CO-CHAIR BASKIN: So, help me understand urodynamics just for a second. What would be an objective demonstration of stress incontinence that is not a urodynamic test? Help me here.

MEMBER PELLETIER-CAMERON: Have them lay down and cough.

CO-CHAIR BASKIN: That is objective?

MEMBER PELLETIER-CAMERON: Yes. You put a speculum in. You say, "Cough." And then, if urine squirts out of the urethra,

that is objective demonstration of --

CO-CHAIR SAIGAL: That is better than urodynamics probably.

MEMBER PELLETIER-CAMERON: -- incontinence, yes.

CO-CHAIR SAIGAL: Okay. So, then, I think we can vote on the overall approval of the concept, yes or no.

MR. WILLIAMSON: You will now vote on the overall approval of the concept. There are two options, yes and no.

You may begin voting now.

(Vote taken.)

And we have 15 yes and zero no.

CO-CHAIR SAIGAL: Great.

MR. AMIN: Chris, can I just jump in here real quick for a second?

CO-CHAIR SAIGAL: Yes.

MR. AMIN: Sorry. I know we are a little behind schedule, and I appreciate the fact that you are on top of it.

One of the components of feedback

that we got from the first few measures that I just want to kind of push back to you guys is that this is the opportunity now, now that stage one has ended and you are going forward as stage two, to review also any concerns that you have about the way the measure is constructed.

While the validity of the measure will be evaluated in stage two, if there are concerns, in essence, of the way the numerators are defined or other concerns that you had, this would also be a place to describe those concerns before the measure is fully specified and you evaluate it in stage two.

So, an example of that would be, if you were looking at, for example, a prior measure that was looking at five different components or a couple of different components, if you agreed that some components were important and others weren't, or if you wanted to expand the numerator, this would be

the place to provide that feedback.

DR. PACE: Another place where you may want to look at is the exclusions, whether the people being excluded, or potentially being excluded from the measure are appropriate, you know, or if you want to make any suggestions or comments about that.

CO-CHAIR SAIGAL: Okay. Any such comments for the developer?

CO-CHAIR BASKIN: Yes. And this is actually a comment that I hate to have to repeat several times today, and it wasn't about exclusions because I have written down for comments about many of these measures this whole idea of excluding people, medical reasons for not performing a workup. And I have a little trouble with this, in that if you have medical reasons for not performing this basic workup, and, yet, you are okay to have surgery done, I am having issues with that.

The other issue I have is this

whole thing with cognitive impairment. If somebody is cognitively-impaired, then there is a medical guardian who is making the decision. I don't understand why that person, this discussion, why this workup still wouldn't have to happen. I mean, there is informed consent by this other person, but I still don't understand why the workup doesn't happen when there is cognitive impairment. Your history may be taken by the medical guardian, but it still obtained.

It just seems to me that this just opens a door that shouldn't be opened. I mean, the 1 or 2 percent of the people for which it is absolutely impossible to ever do this, I guess is one thing, but I just don't see it happening. And I think these are just exclusions that become shady and subjective and just shouldn't be allowed. They weaken the measure.

CO-CHAIR SAIGAL: So, the developers could take home the idea that they

would have to have a better justification for those exclusions.

Okay. So, can we move on? We have three more of these to do in 30 minutes.

So, C2050, Alayne, could you lead us? This is the counseling-on-treatment options and importance to measure and for high impact.

MEMBER MARKLAND: I am going to do a brief overview, and then state my findings in this review.

Briefly, this is the percentage of female patients who are having stress urinary incontinence and surgery, that a documentation that treatment options were discussed. These treatment options are twofold. Treatment options include non-surgical therapies, both behavioral, which is defined here, and other surgical treatments.

And then, the second part of this is expectations for treatment. What are the rates of potential cure or potential problems

with the complications of having surgery?

And so, it is the twofold measure. The numerator statement is as such. The denominator is excluding female patients who had stress surgery without concomitant surgery for pelvic organ prolapse. So, these are just women presenting with one type of surgical intervention.

And then, the same denominator exclusions, in addition, are the medical reasons for not counseling patients, including the cognitive impairment with the statement as just said, and those patients who might be uncomfortable with the responsibility of making choices regarding their care.

So, with this, the timeframe is a 12-month period. So, the discussion, I guess, on options is specified as 12 months. They do list the types of surgeries here -- I am not going to read those out -- and the types of different behavioral or non-surgical treatments, including non-pharmacologic

measures, and the actual rate in the literature in terms of cure rates that they are stating here for each of those surgical measures that they are trying to look at what kind of treatment options were discussed.

The level of analysis is both administrative claims and paper medical records. Really, the level of analysis is at the clinician or the individual level in terms of these discussions.

And the impact is that really large amounts of surgeries, as we have already heard, are being done for stress urinary incontinence. This is a really fairly easy procedure with little time for hospitalization, if at all, commonly being done. Stress incontinence is a broadly-accepted problem among older females as well as younger females.

These surgeries can be very effective at improving outcomes. However, oftentimes, they state the evidence that

really people don't counsel about other types of treatments in the broad sense, maybe not in specialty groups, but maybe more providers in different areas.

And sometimes women come to have surgery without having all other treatments informed of them. We don't have data saying what that number is per se, but it is fairly common. And the authors do cite here that 97 percent of the women who have surgery really expect this to have been the cure, and that is often not the case. The cure rates are more in, what they report here is 69 to 82 percent with some variability.

CO-CHAIR SAIGAL: So, then, in terms of the impact, this is counseling people appropriately for a high-volume surgery is a high-impact measure, is that right? Okay.

Anyone else on the Committee have any comments about that?

(No response.)

Okay. So, can we move on to the

impact part of this then?

MR. WILLIAMSON: Okay. We will now vote on the impact. There are four options: high, moderate, low, or insufficient.

You may begin voting now.

(Vote taken.)

CO-CHAIR SAIGAL: Okay.

MR. WILLIAMSON: And we have 12 high, 3 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Great. Okay. So, then, maybe talk about the evidence supporting their measure?

MEMBER MARKLAND: The evidence cited with this guideline includes statements or guidelines from European societies as well as U.S. societies. And basically, they list options here, that non-surgical treatments can be very effective and are often grade A types of evidence for treatment for this problem and should be considered in the treatment line.

Let's see, basically, patients should be given all these options in a clinical setting prior to proceeding straight to surgery. And I think that is what this measure is trying to establish. But what we don't know is really what are the rates of this discussion happening.

CO-CHAIR SAIGAL: So, to summarize what you said, basically, there is evidence, they provided evidence that incontinence is an important problem, but this specific measure about appropriate counseling, there is not a direct -- there is no evidence that counseling someone completely will improve their outcomes. It maybe isn't a study we can really do again conceptually. It is more this may fall under an issue that we have -- it is a consensus that you should counsel patients appropriately.

And so, any of the other reviewers who read this document have anything to say about the evidence, or other Committee

members?

(No response.)

Okay. So, I think we are going to have to make a vote then. I have feeling this is, again, we already said it is important. The evidence that suggests that it is important in terms of the outcomes, that we have a gestalt that it is important, doesn't really rise to the level of an NQF situations of acceptability.

So, we have the options of saying, yes, we believe it does; no, we believe it does not definitely, or that it does not, but we have a feeling that this is important enough as a problem that we should make an exception.

I mean, we are going to have this discussion several times today. So, I will tell you what the NQF staff told me, that we are not supposed to be doing this routinely, but only in situations in which we feel that the problem is important enough. So, it is up

to us to decide.

MR. AMIN: Chris, can I just clarify, because I think we will be going down this road a number of times? I just want to clarify, No. 2 is where you would be invoking the exception, where there is no empirical evidence here. It is okay. I just want to make sure that we are clear. From the tenor of the conversation, that seems to be where you are going.

And No. 3 would be that the information exists, but that information that exists actually would meet the quality, quantity, consistency requirement, but it just wasn't presented by the developer. And the group would have a discussion around what that evidence entails.

CO-CHAIR SAIGAL: Great. Okay. Thanks for that clarification.

Liliana?

MEMBER BORDEIANOU: Sorry, I am waiting for the green light to go on.

But, then, okay, the clinician in me says, so let's say I am taking care of a patient like that and I document that I have discussed the risks of a surgical procedure and I have discussed all the other options, but the patient decided to proceed with surgery. Does that make me a quality surgeon, just because I documented that in one sentence? Or do we really want to look at how much care was provided to these patients before they actually proceeded to surgery, in which case I don't think that this measurement would measure it.

CO-CHAIR SAIGAL: You say "how much care". What do you mean by that?

MEMBER BORDEIANOU: In the medical record in general, not from the documented note from this particular provider, but the physical therapist that might have taken care of this patient, the nurse practitioners, the primary care physicians. The medical record of a patient is larger than the encounter with

a surgeon.

So, if the surgeon says, "This patient came to me and they already had biofeedback, and I am offering them surgery" --

CO-CHAIR SAIGAL: So, you are suggesting that it would be more convincing if they took into account what other providers the patient had seen?

MEMBER BORDEIANOU: I think that it is easy to cheat the system and say in every note that you dictate automatically, "I have provided counseling," et cetera, et cetera, "and the patient still wants surgery."

CO-CHAIR SAIGAL: I see. So, that is a good question for, I think, the validity, and so on, of the measure. But I think, first, we decide whether if in a perfect world it was valid, would it be a good measure?

And then, when they come back in stage two, they can tell us when they actually tested these things out whether it was valid.

As you point out, you can easily game the system.

So, let's put that aside for right now and imagine it was valid, I think is the idea. Is that right?

DR. PACE: Right, but I think your question kind of also relates to evidence. I mean, because, again, is there evidence that counseling or providing this list of options is the right thing? Or are you saying it should be that they have actually tried some of these lower-level or lower-invasive -- I think that is maybe what you are asking at this stage, is: where is the evidence? Where would the evidence be most appropriate?

CO-CHAIR SAIGAL: Right. So, there is no evidence that they have presented that provides this link. But, as we talked about last time in the last measure, it is pretty standard of care to fully counsel patients. So, will we ever develop evidence around that or should we?

CO-CHAIR BASKIN: Actually, I don't think this evidence was presented here. But I think there is a body of evidence out there that, in general, not related to this particular surgery, that when you provide members with, when you provide patients with treatment options, that people do change their treatment. There is a whole body of evidence about preference-sensitive surgeries and things like that.

So, that I don't know that there is any need to have that evidence specific to this particular surgery and for every surgery that is contemplated out there, but I think there is a great body of evidence that decision-support information for patients is a good thing in making an informed decision, and it does change treatment choices.

CO-CHAIR SAIGAL: In that regard, I think if you look at the data on shared decisionmaking, that for sure providing people with all their options does impact their

decisionmaking in general. So, if we were to accept those kind of data which aren't done in incontinence, but are done in BPH or CABG surgery, and so on, if those data are acceptable to the Committee, then we could vote a three, that there is data about decisionmaking that exists that would be convincing enough that it would work in incontinence surgery, but it wasn't presented in the document. So, that is another option we have.

MEMBER PELLETIER-CAMERON: Yes, and, I mean, there is evidence presented in this document, evidence about the success of each of these interventions. I mean, there is good data on the success of pelvic floor physical therapy, good evidence on weight loss, good evidence on all the surgeries. So, there is evidence that each of the suggestions that you can discuss are effective.

CO-CHAIR SAIGAL: Okay. Okay. So, then, to summarize where we are at, we are

going to take a vote on evidence here supporting this measure. This is not direct evidence about the specific measure and outcome that matters to patients that is presented by the developer. This is definitely a standard of care that we all probably, as physicians, feel is important to do. So, we have that option of, if we don't feel evidence is there, we can make an override.

And then, I think there is also an argument that there is a large body of data about shared decisionmaking that says, when you fully counsel patients about any treatment choice, that you tend to get more appropriate treatment choices and you impact the patterns of care.

So, our choices are one, two, or three. Let's vote.

MR. WILLIAMSON: We will now on the evidence.

You may begin voting now.

(Vote taken.)

And we have 3, yes, that the body of evidence meets the guidance. We have 3, no, the evidence does not meet the guidance, and we have 9 that, no, insufficient information was submitted to rate.

DR. PACE: So, the next question is, if we go with this question, if people want to address, to specifically vote that, yes, a body of evidence does exist, just to have that on the record that everyone agrees.

CO-CHAIR SAIGAL: Okay. So, for the plurality of us who voted for three, we, I guess, believe that there is evidence out there, a large body of evidence about shared decisionmaking. So, do you want to affirm that as guidance from the Committee to the developer? Or, if you don't believe that that is relevant, which it may not be, then you can vote no.

Zahid?

MEMBER BUTT: When we say there is

a body of evidence that exists, do we need to also provide the reference?

CO-CHAIR SAIGAL: No.

MEMBER BUTT: Or is it just simply to acknowledge that it exists?

CO-CHAIR SAIGAL: Yes.

MEMBER BUTT: Okay.

CO-CHAIR SAIGAL: Yes, it is specifically, you know, the Foundation for Informed Medical Decisionmaking has many studies on this. Al Mulley is one of the authors, Mike Barry.

Okay. Anything else?

(No response.)

We should vote. Let's vote.

MR. WILLIAMSON: This will be a hand vote, and I believe these slides are not working. So, we will now vote if there is general agreement that the quantity, quality, consistency of the body of evidence meets the NQF guidance.

So, please raise your hand to

indicate yes.

(Show of hands.)

And raise your hand to indicate no.

(No response.)

It is unanimous. We have 15 yes, zero no.

CO-CHAIR SAIGAL: Okay. Then, Alayne, the last comment is on performance gap.

MEMBER MARKLAND: The last comment being I do think there are some people who may do a part of this, but not all. Specifically, the counseling on the procedure itself, maybe not always about other treatments. So, I think there could be a performance gap that could be measured here in this measure.

CO-CHAIR SAIGAL: Okay. And they list some data about that under 1(b)(3). Did you have any comments about that body of data?

MEMBER MARKLAND: Yes, I think some of this data, though, that they list here

isn't as applicable to a U.S. population. We don't often cover pads and discussion of pads, although the VA does, but in the U.S.

And so, I think this discussion could be improved on in the data that they list here. But what they do list here is what patients' expectations are. And I think that does factor into this equation, that often patients really have high expectations for surgery and may not realize other treatments may be also helpful.

CO-CHAIR SAIGAL: Okay. Any other comments about the performance gap?

(No response.)

Okay. So, it looks to me like there is some evidence, some observational data that there are patterns-of-care variations that sound totally believable to me. And I do believe I am sure that not all doctors provide all the appropriate counseling for patients before they do surgery.

So, let's vote. Okay. Let's

vote.

MR. WILLIAMSON: We will now vote on the performance gap. There are four options: high, moderate, low, or insufficient.

And you will begin now.

(Vote taken.)

And we have 3 high, 11 moderate, 1 low, and zero insufficient.

CO-CHAIR SAIGAL: Our final overall vote now -- do you have questions? Okay. So, to summarize this concept, which is basically looking at appropriate treatment counseling prior to surgery for stress incontinence, we felt that it was an important thing to tackle in terms of the number of surgeries that get done and the variety of options women face.

We felt that the evidence surrounding the measure was generally indirect and provided from things like literature that this would be a good thing to do.

And in terms of the performance gap, there is probably moderate evidence that there is a gap, but we believe that it is sufficient to move forward with it.

And then, we can vote about whether we want to approve this concept overall to provide to stage two.

MR. WILLIAMSON: We will now vote on the overall recommendation of this concept.

You may begin voting now.

(Vote taken.)

And we have 15 yes and zero no.

CO-CHAIR SAIGAL: Great.

Okay. So, then, we have two more to do

MEMBER FALLER: I was just going to echo your comment on the last one about the exclusion of informing people who had cognitive deficits that somebody would be involved in the decisionmaking. And just to do the surgery without counseling somebody doesn't --

CO-CHAIR BASKIN: And even in addition, this has an additional one about people who have patient reasons for not counseling and this whole thing about uncomfortable. I can't buy that one, and I don't think it will get bought on the next level of review someplace.

I also, though, have a question about why -- and once again, my urology friends may be able to answer this for me, and it wouldn't go any further than this suggestion -- this whole thing about continent and prolapse. I don't understand. If they are also getting urinary incontinence surgery as a part of their surgery, why would you exclude those people? The whole idea of treatment options and counseling people still exists. I don't think there is any specific reason why you have to have a pure denominator group of homogenous people who are getting strictly urinary incontinence surgery without prolapse surgery at the same time. Why this

measure wouldn't apply to all those folks at the same time?

CO-CHAIR SAIGAL: Go ahead, Stu, please.

MEMBER REYNOLDS: Well, it is certainly a complicating issue. Those treatments, while they may be concurrent in some people, are very different and have their own logistics, including risks/benefits, and whatnot.

And even going back to the last one, the work up there is also a little bit different in the assessment of the degree of prolapse. And so, the conditions can exist in the same individual. That individual could choose to have one or both of those things repaired, but they are also very different. And so, I feel pretty comfortable that they split this out, because it is a subset population, that someone who has prolapse is sort of a different ball game in terms of how you are going to counsel them, the workup you

are going to do, and all that kind of stuff.

CO-CHAIR BASKIN: But the measure has nothing to do with workup. The measure is simply that you have counseled them on treatment options. So, those folks still have treatment options and still require counseling. Even though they may require additional surgery and additional testing, that is really not what the measure is measuring. It is measuring whether you provided treatment options and discussed them.

MEMBER REYNOLDS: That is a fair enough comment. I was lumping, also, the last one in as well. And you sort of had brought up that with many of these ones. But, certainly, the argument may be less strong with this specific measure --

CO-CHAIR BASKIN: Yes.

MEMBER REYNOLDS: -- than the assessment one, which would be different.

CO-CHAIR SAIGAL: I think, Stu, basically, with prolapse, I mean, you can have

other treatment options as part of your recommending them. And so, maybe you wouldn't just say you are a good candidate for Kegels if you have got bad prolapse. So, it may be a different cohort to counsel. So, they wanted to make this a homogenous enough of a group that the treatment options that they were counseling about were appropriate for this group.

But in terms of the other exclusions, I agree with you. I think that the developers should reconsider the other denominator exclusions and the rationale for them, because I agree with you on that.

Zahid, do you have a question?

MEMBER BUTT: Yes, just another comment. I made a comment earlier about encouraging developers to do e-measures, but I also want to make a comment -- and this may be sort of for all the measures -- the data source here is specified only as claims data and paper records, but it should also include

EHR data because, even though someone is answering the CPT question, they could use an EHR to look at the information. This one, just for completeness' sake, limits to only paper record.

CO-CHAIR SAIGAL: Okay.

CO-CHAIR BASKIN: And one last comment. So, if this does remain as-is with just strictly urinary incontinence patients, then I can't imagine why this and the prior measure wouldn't be combined as one measure. Every patient needs assessment. Every patient needs treatment options. And essentially, it just calls out for a composite measure that says you have done both of these things. Any less than both of these things doesn't make any sense.

CO-CHAIR SAIGAL: Okay. Any other comments from developers?

(No response.)

No? Okay.

So, then, the next one is 2051,

patients counseled about risks associated with the use of mesh in sling surgery.

And Dr. Ellis is going to do that.

MEMBER ELLIS: So, remember everything you just heard about the last measure because a lot of it just transfers over as we keep subsetting this group down.

(Laughter.)

So, now we are talking about the patients counseled about risks associated with use of mesh in sling surgery prior to surgery. The author presents reasonable evidence that a majority of that growth in SUI surgery we are seeing is being driven with mesh use.

The impact to that group, if we kind of skip up to a little bit of the evidence part, there is a lot of European evidence cited here showing a lot of adverse outcomes and complication rates and the like with mesh.

There is less reference to it on the American side. There is reference to an

FDA alert. There is some reference to the construct of that, I use the word "counsel" because that is the measure word, although in the European citations it is almost always "warn, warn, warn," not "counsel, counsel, counsel". So, you have to kind of make a distinction in how those two words are used.

But, essentially, the population at risk here I think has been defined by these kind of previous measures. That kind of defines the impact group, if we want to start there.

CO-CHAIR SAIGAL: Okay. So, the impact here is a very common surgery. This material being used has been documented to have a specific risk that is different from other materials used to suspend the urethra. This is a high enough impact problem that it should be measured at the national level. That is the question before us.

Any comments about that?

MEMBER ELLIS: I will say, for

guidance for some of you, in the preliminary evaluation it was kind of split on high and medium from the original reviewers.

CO-CHAIR SAIGAL: Yes. Okay.

Want to vote?

MR. WILLIAMSON: We will now vote on the impact. You have four options: high, moderate, low, or insufficient.

And you may begin voting now.

(Vote taken.)

And we have 9 high, 3 moderate, 2 low, and 1 insufficient.

CO-CHAIR SAIGAL: Okay. So, we can move on to the level of evidence supporting the measure.

MEMBER ELLIS: Again, the citations, a lot of citations specifically about outcomes using mesh, a lot of the stuff out of European studies, a lot of RCTs cited out of Europe, very little out of America.

And in neither case were these studies about counsel. It was all about

outcomes related to use of mesh. So, we have to kind of take that same leap of behavioral change relative to being engaged and informed as to the risks associated with it based on the results of those outcome trials as opposed to the outcomes of counsel.

There was more than enough evidence on the European side, I think, submitted to support the notion that these risks are increased with this material, less so, although we do have an alert from the FDA on the American side.

CO-CHAIR SAIGAL: Right. So, mainly, it is observational data supporting it. There is a reasonable quantity, according to NQF standards, of European and American studies.

I personally think that the FDA action is significant in terms of its relevance for the U.S. population.

Any others? John?

MEMBER MORTON: I was just going

to ask, for someone who is not super-familiar with the slings, what was the source of concern from the FDA? Was it erosion, sort of inappropriate use, putting them in patients who didn't need the mesh?

CO-CHAIR SAIGAL: It was mainly erosions.

Okay. So, if no one has any other comments on evidence, we can vote that it meets guidance, doesn't meet guidance, or there is data we are aware of that has not been presented.

CO-CHAIR BASKIN: Sorry. I am not quick enough raising my hand here.

This is unlike the previous measures in that where there is a discussion of different options. This is somebody has got a planned surgery, and have they explained the risks of the surgery or have not explained the risks of the surgery. I don't understand how any of the evidence here is anything about whether explaining risks to people of a

surgery that has already been selected, as opposed to making a choice of here's two different ones, a medical treatment versus a surgical treatment. That is a very different thing.

This is a check-the-box measure that just says, "Check the box that you gave somebody informed consent," which I am presuming is required. I don't even see this as a measure, let alone that there is evidence to support what we are measuring here. I mean, I don't understand. Are we saying that people are going to not have the surgery because they weren't explained risks before and now they are going to be explained risks, and the people are going to change their mind about having the surgery? I don't see any evidence that is pointing to that.

CO-CHAIR SAIGAL: John, go ahead.

MEMBER MORTON: Well, I think the one thing that makes it fairly unique -- I agree with you, this falls under informed

consent. You go over risks, benefits, go through that equation; the benefits should outweigh the risks.

But I think what makes it unique is this issue from the FDA that makes it a fairly unique problem. Rather than waiting until a lot of data comes out, if there is this FDA ruling, there needs to be a specific recommendation around it. That is my impression.

CO-CHAIR SAIGAL: I think you are making a good point. This is similar to the last one, though, in that there is not a direct -- I mean, I think it is important to mention mesh because it would stop a patient from having mesh if they mention they want to use it, if they told them all about it.

But the idea that you would counsel them and then they would have less mesh erosion, there is no direct evidence about that. So, in that sense, it doesn't meet the NQF evidence guideline.

I think it might meet the guideline if we consider the other evidence on shared decisionmaking, like we did the last time; that when you fully inform patients, they can take information and make different decisions.

MEMBER ELLIS: I think the problem there with that, of course, is that both the numerator and the denominator are people who had mesh sling surgery, not an indicator that we had a change in choice, right, to a different type of surgical intervention.

So, I think that is what muddies the water in terms of the informed consent part of this, unless we make a leap that informed consent truly affected the outcome of that mesh surgery. And I think that is not a leap we are willing to take.

CO-CHAIR SAIGAL: So, that would argue maybe to change the denominator to include all patients that are being counseled for surgery that have a mesh -- Jenifer?

MEMBER LIGHTDALE: It is informed consent versus shared decisionmaking. I mean, it is difficult to call it shared decisionmaking if the moment of discussing risks is happening at the time of the procedure. I think that is very different. So, I think you have a --

CO-CHAIR SAIGAL: Well, it is before the procedure, I would think, yes.

MEMBER LIGHTDALE: Right, right, but immediately before a patient is NPO, prepared themselves psychologically to undergo a procedure. It is a very different experience. Counseling them at that moment about an FDA warning is very different from --

CO-CHAIR SAIGAL: Is that what it says, right before the surgery?

MEMBER ELLIS: It just says before.

MEMBER LIGHTDALE: No. Well, it just says before. I mean, talking before, it could be at that moment of informed consent,

right?

CO-CHAIR SAIGAL: So, you are saying that, to make it of high impact, it would be done not in the preop holding area, obviously, which I think makes good sense, right?

MEMBER MERGUERIAN: May I maybe suggest another concept? Maybe for the AUA to actually have an informed consent written up with all the pros and cons of these types of surgeries. And then, the question would be, was that actually given to the patient, yes or no?

CO-CHAIR SAIGAL: Okay. So, then, in terms of the evidence, we have evidence that there is a problem with mesh in this country. We have the measure that says, if you tell people about mesh, that is a good thing to do because it will decrease, the implication is that it would decrease mesh complications, I guess is the idea. No? I mean, that is how I guess we could measure it.

That is a patient impact.

MEMBER ELLIS: It seems like the only way you would get there the way this measure is presented.

CO-CHAIR SAIGAL: Yes. Right. That is a stumbling block conceptually. Because what you are saying is, potentially, the only way to avoid mesh complications is not to use mesh, and then you could drive utilization to zero. That is the idea that I am reading from this measure.

Yes, Ed?

MEMBER GILL: We are supposed to be, as far as I understand it, evaluating it based on the measure as it is written. And we are sort of trying to read into all of these things and put our own spin on it. But, I mean, basically, all they are measuring is was the patient counseled appropriately that had the surgery. And everything else is sort of secondary to that, I think. I mean, we are really just trying to evaluate whether people

are being counseled appropriately for their procedure.

MEMBER ELLIS: Yes, I think that is a big leap, is the appropriate part. There is a lot of variation, if you look through their evidence, on what one might assume is appropriate.

CO-CHAIR SAIGAL: And the evidence the NQF is asking us to look at is evidence that links it to an outcome that matters to patients. So, the outcome that matters to patients in this setting would be, it could be either I didn't have a complication from mesh or I had a complication, but I was aware there was a risk. One of those two things is the outcome, I guess.

Stu, you had a comment?

MEMBER REYNOLDS: I was going to say one other outcome could be whether they chose to have the operation or not.

CO-CHAIR SAIGAL: Right.

MEMBER REYNOLDS: I don't know if

you said that or not, but --

CO-CHAIR BASKIN: But not in this measure because the measure is only, as Robert says, it is only those patients who had the surgery.

MEMBER REYNOLDS: So, my second comment is it is tempting to look at this one in the framework with the other ones that are all sort of part of this. I guess a question I would pose to everyone is, do you kind of go forth with it and say, well, maybe at the end we will kind of condense into one on counseling and discussion, or do we take it as it is, a standalone and say, is it good enough to go through? And I don't know the --

MR. AMIN: Chris, let me jump in here --

CO-CHAIR SAIGAL: Go ahead.

MR. AMIN: -- on some procedural options that you have. So, what we talked about doing is having a harmonization discussion. Essentially, what that

harmonization discussion will entail is precisely that. While you want precisely that, you may make some recommendations about, first, if there are components that seek to measure the same care process and should be combined or that there is some logical harmonization related to the population being measured.

So, what you should do right now is evaluate this individual measure as it is constructed. If you feel that you want to, then, use this measure and recommend that this measure should be combined with others in the future, in the future conversation we have later this afternoon, that would be appropriate at that time.

So, I guess that your procedural options, I guess.

CO-CHAIR SAIGAL: Zahid?

MEMBER BUTT: Just one more comment, that this might be the type of measure where experience abroad might be

relevant because it is really a complication of something that is put in. One would assume that it is put in correctly by different people, as opposed to more practice patterns and what you do in certain situations. Here something has been done, and it erodes through.

CO-CHAIR SAIGAL: Yes.

MEMBER BUTT: And if it is eroding through in Australia, it might be relevant in the U.S.

CO-CHAIR SAIGAL: I agree with that. But I think the issue before us is really, I think we all agree that the evidence is strong that there is a problem in this country. The problem is how the NQF wants us to measure this evidence supported in a measure. Does it support the measure as written, which basically looking at people that have mesh surgery and have never been counseled around the risks of mesh.

So, I am just not clear what the

outcome is. The outcome, I guess, implied is that you are aware of the risks of mesh and you had surgery. That is just the outcome.

MEMBER BUTT: I think it really just has to be what was mentioned earlier, that somebody might choose not to get it done.

CO-CHAIR BASKIN: That is not the measure, though.

CO-CHAIR SAIGAL: Yes.

CO-CHAIR BASKIN: This is not counseling about treatment options. There is no option here. You have had the surgery. It is a lookback to see whether somebody told you you could have had a risk. This is not measuring whether you changed your mind.

MEMBER BUTT: But if, let's say, that if you counseled your patients regarding the three specific things that are in the measure numerator, erosion, exclusion, pain, permanence, and you counseled it in 100 prospects that you were contemplating surgery on versus only 10 percent, I would imagine

that the ones --

CO-CHAIR SAIGAL: But the denominator is people who have had mesh. You only identify people who have had the surgery. The people that didn't have the surgery are not captured.

MEMBER BUTT: I understand. I understand. But it still gets to the provider's practice pattern, whether they counsel 100 percent of the ones they did or only 10 percent. And I would imagine that, if they did that as part of their engagement with the patient, that the one that only mentions this 10 percent of the time will probably haves more people willing to go through it than the ones who tell this complication to 100 percent of their people that --

CO-CHAIR SAIGAL: So, the evidence is even more tenuous. That may be true.

MEMBER BUTT: I am just saying that --

CO-CHAIR SAIGAL: Yes.

MEMBER BUTT: -- is kind of the only sort of --

CO-CHAIR SAIGAL: Link.

MEMBER BUTT: -- way you can link this as to where there may be a difference in outcome.

CO-CHAIR SAIGAL: Right. Yes.

All right.

MEMBER PELLETIER-CAMERON: Just for my own clarification, so right now we are kind of at the stage where we can either accept or reject this measure. But, from the discussion, I hear a lot of people agree that this is important, but that it almost seems like it needs to be harmonized with the previous one where seems to fit in much more appropriately.

So, I can tell you what I think is I think it should be harmonized with the previously one because your denominator and numerator now make sense. And what direction do we go in to have that happen?

MR. AMIN: So, then, you would move this measure forward. And then, we will have a discussion at, I think it is at three o'clock, where that would be the recommendation that you put forward.

CO-CHAIR SAIGAL: You could not move it forward and still recommend it to be harmonized into a measure later. It doesn't have to move forward to make a recommendation of harmonization.

MS. WILBON: So, the other option is you don't have to move the measure forward, but we could add a note, like we have been doing for every measure, your recommendations to the developer, that we will be giving each developer a checklist to say the Committee wants to see you do A, B, C, and D before it comes back to stage two. That would be an addition that we would make to one of the prior measures, to say we want you to add this to the numerator before you bring it back to stage two. That is your other option.

CO-CHAIR SAIGAL: Okay.

MEMBER MORTON: Harmonization does not require prior approval of the measure? Do we have to approve this measure, so it can be harmonized later?

CO-CHAIR BASKIN: No, no, they would have to bring it back as a measure or bring it back as a harmonized measure or something like that.

MS. WILBON: Right. When it came back to stage two, we are still working this out, but there would be a period where we kind of review what you recommended, what they brought back, does it match. Does the evidence kind of match what they actually brought back? So, that is what we are thinking.

MEMBER ELLIS: I just want to make one last comment on Zahid's comments about using it kind of in its current form as an evaluator of physician performance, for example. Even though there is very little

reference to that concept in this measure, it also struck me as troublesome, if we tried to take it down that path, and it being, basically, an administrative measure. You know, check the box. "Yes, I did," right? I mean, relatively easy thing to do with no construct of what is quality counsel, what were the options that were required. So, I think that is a dangerous way to take it as well in its current form.

CO-CHAIR SAIGAL: Okay. So, I think we should probably move to a vote on this. It sounds like there are a lot of conceptual problems some of us have with this measure. The denominator is only people that have had mesh surgery, which limits its potential usefulness as a measure.

There is a recognition that there is evidence that the mesh is bad, but the evidence that this measure will impact things that matter to patients is limited from the NQF standard in that it not even meet the

issues we raised previously on shared decisionmaking because of the denominator issue.

So, we can vote on this. We can turn it down and ask that it be harmonized. We can vote it up and ask that it be harmonized. There are both options for us. Okay?

Let's vote.

MS. WILBON: So, can I just make a point of process? This vote is still on whether the evidence submitted before you meets the criteria. So, I know there has been a lot of discussion about whether or not the concept should move forward, but we are kind of still back to the foundational information that was submitted on whether or not what they submitted supports the measure focus and based on that.

CO-CHAIR SAIGAL: Okay.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

We have 5, yes, that the body of evidence meets the guidance. We have 7, no, that the evidence does not meet the guidance. And we have 3, no, that there is insufficient information submitted to rate the evidence.

CO-CHAIR SAIGAL: So, for those who voted three, are there any specific studies that you are aware of that you -- that is the implication there, I think, is that there are other data out there to support the measure focus that wasn't submitted here.

Anyone who wants to comment about that? Is that the idea that you had about that? Any ideas?

(No response.)

No? Okay.

So, can we proceed?

MR. AMIN: Here, well, you are going to have to have a discussion around -- voting two would essentially bring us back to

where we have been, which is invoking the exception.

CO-CHAIR SAIGAL: Okay.

MR. AMIN: So, your discussion around invoking the exception -- actually, Evan, can you move to the next slide on the exception.

CO-CHAIR SAIGAL: Okay. So, now we have got to vote as to whether, although the majority of people felt or a lot of people felt that the measure did not meet the evidence criteria of the NQF, is this a big enough problem? And do we think that there is a fundamental relationship between the measure and the outcome that matters to patients, that we would vote to make an exception here and let the measure move forward? So, let's vote on that.

Is the voting open?

MR. WILLIAMSON: No, I didn't open it.

CO-CHAIR SAIGAL: Okay.

MR. WILLIAMSON: This will be a hand vote again. This will be a hand vote.

So, first, raise your hand for yes.

(Show of hands.)

All right, so we have 3 yes, 4. All right.

And for no?

(Show of hands.)

MR. AMIN: That was 4 yes, 11 no. Okay. So, the measure does not go forward, and we can move on to the next measure.

CO-CHAIR SAIGAL: Okay. The next measure is C2052, reduction of complications through the use of cystoscopy during surgery for incontinence.

Dr. Gill?

MEMBER GILL: So, we have talked about this a lot. It has been well-presented.

Stress incontinence is a common problem. The procedure is very common. And so, there is certainly a high impact here, in

my opinion.

And cystoscopy has been recommended by -- three major guidelines recommend this measure.

I guess we will go in order here, though. You just want --

CO-CHAIR SAIGAL: The importance.

MEMBER GILL: -- impact and importance.

So, yes, I mean, I am not going to go through it all again, but it is very important, I would say.

CO-CHAIR SAIGAL: Okay. So, and then it comes to the panel about importance. Again, this is a high-volume surgery. Cystoscopy is used to decrease complications for it. Is that important as a measure nationally? That is the question.

So, we should vote.

MR. WILLIAMSON: We will now vote on the impact. There are four options: high, moderate, low, or insufficient. You may begin

voting now.

(Vote taken.)

And we have 13 high, 2 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. So, now the evidence surrounding this measure.

MEMBER PELLETIER-CAMERON: So, the evidence is mostly the three recommendations from the European Urology, AUA, and ACOG, who have all recommended doing it. It is all expert opinion and consensus. They didn't really present a lot of or any evidence other than those guidelines.

CO-CHAIR SAIGAL: So, there is expert opinion. Isn't there some observational data they mention in this, about this?

MEMBER GILL: Right. I'm sorry, yes. Yes, yes, yes. This one, there is actually a fair amount --

CO-CHAIR SAIGAL: Right.

MEMBER GILL: -- of observational

data. There are some randomized controlled trials as well.

I was thinking of the other one. I'm sorry. There is actually quite a bit of information supporting this.

CO-CHAIR SAIGAL: It sounds like it is an area of controversy to some degree, but that there is at least observational data and there is some CEAs that have been done on this topic.

I don't know, Stu or Anne, if you have any comments about that or you are aware of --

MEMBER BUTT: You are talking about 052?

CO-CHAIR SAIGAL: 2052, using cystoscopy to reduce complications of stress incontinence.

MEMBER BUTT: Yes. I'm sorry.

MEMBER REYNOLDS: What I would say is that, the way that the form is filled out is that there is very little data actually in

the data part, and most of it is back under the impact and the gap, which I think does reinforce or suggest that there is some evidence there, but it is poorly filled out.

CO-CHAIR SAIGAL: Uh-hum. Okay.

MEMBER GILL: Really, to go back to what I said initially, the only evidence they presented were those three guidelines in the evidence section.

CO-CHAIR SAIGAL: Right. So, I think there is a summary under 1(8)(3). They talk about a study by Beckett in the use of cystoscopy and what they found there in terms of complications. There is a TVT study. Those are both observational studies.

What else is in here? So, there is at least two observational data and then there are guidelines.

MEMBER REYNOLDS: In Section 1(b)(2), there are some very specific on interoperative cystoscopy and injuries.

CO-CHAIR SAIGAL: 1(b)(2)?

MEMBER REYNOLDS: That is under the gap.

CO-CHAIR SAIGAL: Right. The gap? Again, this is this TVT study and a study by Gill.

So, without cystoscopy, talking about Gill, only 12 percent of injuries to the lower tract were found at the time of surgery. So, there are at least three observational studies that they cite that show that there is a benefit to patients for doing cystoscopy. Cystoscopy is a two-minute, low-risk procedure, and that is where we are at in terms of evidence. And guidelines, of course, consensus.

MEMBER MORTON: I would like to hear from the urologists, you know, the utility for this. It sounds like, for us in GI surgery, we connect two pieces of bowel. We do a leak test. This is essentially what you are doing with the cystoscopy, make sure there is not a leak. For us, for leak test,

there is some variation. Some people do them; some people don't.

Is there a general consensus this is important to do?

CO-CHAIR SAIGAL: Stu?

MEMBER REYNOLDS: It is analogous, but, I mean, you are not looking for leak. Usually, what you are looking for is actual injury or perforation into the bladder. So, for example, when you do these procedures, one of the risks, that you could directly penetrate into the bladder, into the urethra. And so, you are looking to see.

And one of the great benefits is that, whatever it acknowledges, if you looked in and you saw that, you would just pull it out, and then you can replace it safely. And then, there is virtually no sequelae to the patient. Whereas, if you did not look and you left that piece of synthetic mesh in the bladder, that is major sequelae. So, it is certainly low-risk, high-yield in that sense.

MEMBER MORTON: This is really primarily stones or repeated infections?

MEMBER GILL: Yes. It can be stones. It can be repeat surgery required.

CO-CHAIR SAIGAL: Blocked ureter.

MEMBER GILL: It can be acquired infection, ureter damage, right. It is a big deal. It can be a very big deal.

MEMBER PELLETIER-CAMERON: And some of the controversy that surrounds this is that the retropubic approach, when you go around the pubic bone, you are going right next to the bladder. I mean, you are this far from the bladder.

There is also a transobturator approach where you go through the thigh and approach the vagina from that direction, where you are much further away from the bladder. And that method was devised to minimize the risk to the bladder. I think that is a little bit more where the controversy arises. I don't think anyone would do a retropubic

approach where they are doing a cysto, but the obturator approach people would theoretically do this without doing a cystoscopy, although the rate of injury to the bladder is still there, it is not zero ever.

MEMBER REYNOLDS: There is a little bit of politics, historic politics, that goes into this. And so, some of these types of procedures which are lumped in together were specifically designed ideally so that you wouldn't have to do a cystoscopy because there was thought that people who may be able to do the procedure were or are not qualified to do cystoscopy or not.

I think that a lot of that maybe has gone away, and most people would agree that cystoscopy is -- or is that not a fair statement, that most people would agree that cystoscopy is worthwhile or not?

CO-CHAIR SAIGAL: Jenifer, you had a comment?

MEMBER LIGHTDALE: Yes, that was

my question. What would keep you from doing this look? It sounds like politics maybe or scope availability. I mean, I don't know. Time? I mean, what else?

MEMBER REYNOLDS: Some of it is training. I mean, I think that, originally, when these were coming around, there was concern that, for example -- and maybe you all can speak more to this -- that maybe some gynecologists weren't trained or credentialed to do the cystoscopy, but they may be credentialed to do the sling. And that is a little bit before my time. And so, I don't want to talk too much about it.

MEMBER GILL: No, I think you are right. I think, historically, there were those issues. I think that is mostly historical now. I think we can distinguish between diagnostic cystoscopy that I think everybody that is doing these surgeries should be able to do at the time versus more operative or higher or advanced cystoscopy

that may be a different animal. But I think everybody that is doing these slings now should be able to do cystoscopy.

CO-CHAIR SAIGAL: Zahid, did you have a comment?

MEMBER BUTT: Yes. It looks to me that what is in 1(b) should be really in 1(c), and there is not much in 1(b) for the gap, which I guess we will get to next.

CO-CHAIR SAIGAL: Okay. Andy?

CO-CHAIR BASKIN: So, once again, the non-surgeon here, I am very troubled by what is in 1(b)(2) here, this one-line statement that says, "It is largely acknowledged that cystoscopy improves the safety, but multiple studies have stated that cystoscopy is not necessary and it is economical to avoid performing the technique." And it cites three studies, and you look down and see the three things. Of course, I haven't read those three articles.

So, by advancing this particular

measure, we are advancing the concept that this is standard of care in this type of surgery. I am a little puzzled. Is this really standard of care in this type of surgery or is this a controversial subject for which there are people that say, no, it is not necessary; it is not appropriate; you can pick and choose; it is not economical; it is not whatever? And therefore, I have trouble. We are not creating standard of care here. That is not our job.

MEMBER PELLETIER-CAMERON: There might be studies that conclude that the rate of interoperative injury is low, and then maybe they take that result and say, well, then you don't need to do it.

But I think, general consensus, all the practice guidelines and people who do this surgery routinely, I don't think anyone would disagree that it should be done. The rate of injury might be low, but the consequences of a missed injury are very high.

So, I think that might be the only difference; the rate might be low.

CO-CHAIR SAIGAL: It is a cost-effectiveness argument that is being made. Like per injury, you avoid your spending a lot of money on cystoscopy. Those three articles I think are talking about those kinds of issues.

CO-CHAIR BASKIN: But that is a reasonable thing. I mean, there are a lot of surgeries and a lot of GI guys can say the same thing. I mean, you know, to what extent do you have to check when checking and doing something extra takes more time, more money, more everything, and the yield is very, very low? Once again, is that standard of care or is it not standard of care?

I am still not so sure that consensus standards should say that, geez, it seems like a really good thing to do because we are going to find injury and it is not that harmful to do, so everyone should do it. I

don't know that that is evidence that says this is the standard, and that somebody would be, you know, it would essentially be malpractice not to perform this procedure.

CO-CHAIR SAIGAL: Well, I don't now how much we consider economic considerations in these measures. I mean, I am not sure what the answer is to that. I mean, I have never done it before.

MR. AMIN: Well, in terms of the high impact, that is certainly where we will look at the evidence. I mean, if the result of this poor practice results in high cost, then that is certainly something that should be considered in high impact.

But just keep in mind, you know, as we walk through this, we talked about impact. Where we are right now is still evidence. And reviewing what is in the evidence component and not necessarily going to what is in the gap component is really the way you want to walk through this.

MEMBER REYNOLDS: Well, I was going to say that, essentially, what you have to weigh is that there are three of the guidelines which are pretty unanimous in their recommendation that interoperative cystoscopy be performed. I would argue that that, then, becomes the standard of care versus those other three articles that you point up which say it may not be as -- they are in different parts of the document. So, it is hard to know how to interpret that. Certainly, in the evidence part of the document, they point to the consensus guidelines which are those three guidelines which would suggest that cystoscopy would be considered routine or at least more standard of care.

CO-CHAIR SAIGAL: John?

MEMBER MORTON: I think anytime you are entering into another procedure like a cystoscopy, you get asked the same question, you know, risk/benefit. So, for us, if we are looking for a leak, we can introduce a

gastroscope, and we might go right through the anastemosis. It sounds like that is not the case here.

Is there some sort of negative to doing the cystoscopy? Could you injure something?

CO-CHAIR SAIGAL: There is a 1 percent chance of a bladder infection. That is basically the issue.

MEMBER MORTON: And I guess my other question would be in regards to the cost. Since you are doing another primary procedure, isn't this kind of bundled in? And so, the incremental cost is actually not substantial? I don't know.

CO-CHAIR SAIGAL: I would think that it is bundled. It is bundled. So, it isn't really a --

CO-CHAIR BASKIN: Bundled in in terms of a separate reimbursement doesn't mean it costs less. The cost of doing the procedure for the facility, taking the scope

out, cleaning the scope, the nurse, the time, and all that kind of stuff, is still there, whether it gets separately reimbursed or bundled in. So, there is a cost to doing anything that takes out another instrument and uses it and takes the time to do it.

CO-CHAIR SAIGAL: Johannes?

MEMBER KOCH: Yes, and just the point, again, it is about the evidence here. We are not grading whether the standard of care is a good idea. We are saying that there is evidence. And it doesn't appear that anybody is making the argument that there is evidence that says you should do this. There is consensus, but there is no evidence.

CO-CHAIR BASKIN: There is evidence in the document. There is observational data that it --

MEMBER KOCH: Observational?

CO-CHAIR BASKIN: Yes.

MEMBER BUTT: Right. It is just not presented in the evidence section. And

so, that is why people may not have seen it.

CO-CHAIR BASKIN: Okay.

MEMBER MERGUERIAN: They actually cite one article about the cost. They cited a 2005 article that they said the cost of complications doubles. It doubles your cost. So, performing a simple cystoscopy may actually reduce that overall long-term cost of complications.

MEMBER MORTON: I guess one more comment. We have ways of grading the evidence. We saw that slide you presented earlier, you know, randomized trials, more than five studies. I mean, based on that, do we have it?

MR. AMIN: Well, keep in mind that, when you are looking at guidelines that are based on consensus, that would not meet the criteria for quality, quantity, and consistency. I think, again, we want to try to keep this as systematic as we walk through this as possible.

It sounds like you need to vote on the evidence component. If you seek to, again, invoke the exception, there needs to be a discussion on how the benefits outweigh the harms here.

But, in the way that it has been discussed, this would not meet the NQF criteria in terms of quality, quantity, and consistency, based on guidelines that are predominantly based on --

CO-CHAIR SAIGAL: But wait a minute. Because, right here, you say that two to four studies is moderate.

MR. AMIN: Two to four, and then --

CO-CHAIR SAIGAL: So, there are at least three studies that are cited here about they are observational that show that there are various documentations of injuries without cystoscopy and with cystoscopy, I think.

MR. AMIN: So, that is the quantity. And again, I am not here to make a

decision for you. If you feel like that meets the criteria, that is fine. The quantity, you have the five studies.

Evan, if you can move to the next slide on the quality, looking at direct evidence of the specific measure focus and adequate size and the precision of the measure. And again, this is information that they should present to you in the evidence component. Moderate is not in RCTs with controls for the confounders, and low -- you can read it for yourself.

But, then, the important thing, also, is here in terms of consistency, in looking that you have clear clinically and practical, meaningful benefits and harms in terms of the direction and magnitude of the benefit. So, it is a high bar. There is a clear high bar here.

CO-CHAIR SAIGAL: Sure. So, in my reading of it -- and maybe I am wrong -- basically, in terms of the quality of the

data, it is low, and the number of studies is moderate, and the consistency I think is consistent.

I mean, the issue that Andy raises is important about resource utilization, but that is not a question as to whether the direction of the benefit is positive or negative. It is a question of whether it is worth the money.

I think it is hard for us to say. I mean, I don't know if we are that kind of a panel to say that there is some financial criteria that the intervention has to pass for us to approve it. I mean, because they are not saying that the study said that this wasn't worth doing because it doesn't find problems. They are just saying that the cost associated with finding those problems is high. So, that is a different dimension of evaluating this. I just don't know at what point we are supposed to do that. I think we shouldn't, frankly.

MR. AMIN: Not in this evidence. I mean, the resource use component would come in under high impact. That's it.

CO-CHAIR SAIGAL: But it is high impact in terms of the -- it is a value question. That is what Andy is raising, is a value. It is the impact over how much it costs to get there, which is different even from high impact, I think.

MR. AMIN: Right. Well, then, it doesn't feel like it has the space right now. I don't know that the cost-effectiveness question --

CO-CHAIR SAIGAL: Yes.

MR. AMIN: It is an important question, but I don't know where you fit it.

CO-CHAIR SAIGAL: Yes.

Andy, what do you think about that?

CO-CHAIR BASKIN: I mean, part of me says that I don't really want to take into account cost-effectiveness. I am just not so

sure that I am just talking cost-effectiveness here. I am talking about making, based on evidence that doesn't tell me that this is absolute standard of care -- that I am measuring something that a reasonable urologist or gynecologist may turn around and say, in this case, it is not appropriate. And I don't know there is enough evidence that tells me that every time this is the appropriate thing to do.

So, it is not just about the extra cost and time. I mean, I am even going to set that aside. That wouldn't be the reason for me to say no on this one. I have just not been convinced from this that there are three sets of consensus standards that are not necessarily totally evidence-based that say this is the only way to do it; I mean, this is the care, and we are going to measure whether people are actually doing it or not. I just don't -- I haven't made that link.

CO-CHAIR SAIGAL: Zahid?

MEMBER BUTT: Yes, I guess this sort of comes back to that original question about where do you put guidelines in because most practicing physicians, when their specialty gives them a guideline, they treat that as standard of care, however they got to that guideline. And so, that might be a very important question, perhaps not for this discussion, but a future discussion, because that is considered, for all practical purposes, the standard of care.

CO-CHAIR SAIGAL: Yes.

Anyone else have any comments?

(No response.)

All right. So, Stu? No? Okay.

So, I mean, maybe I am going to try to summarize this very complicated discussion. So, this is what I think I heard from the group.

In terms of the evidence supporting this measure, that there is low quality of evidence, some observational data

that is not well-controlled that indicates that cystoscopy has a benefit. And there is also a consensus from guidelines that it is a good thing to do. And the consensus statement stuff, we don't meet the NQF quality criteria.

There is a moderate quantity of evidence, and I think the direction of the evidence is in a positive direction for cystoscopy.

We had a side discussion about whether it is worth the cost because there were data that the developers cited that some people thought it wasn't cost-effective. But, at this point, we are not going to include cost-effectiveness in our decision.

And some of us noted that the evidence was presented in the document in a variety of places where it could be better organized.

So, I think, based on that, we should vote as to whether we feel that the evidence meets the guidance for support,

whether it does not meet the guidance for support, or whether there is other data available that we are aware of that would support the data that wasn't presented here. Okay.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

CO-CHAIR SAIGAL: Okay. So, in this case it did not meet the evidence basis. So, we have got to vote as to whether this is an important enough measure that we would go around the NQF standards for data to proceed. And that would be on the basis that we think this is just the gestalt or the reading of the measure, it sounds like a good thing to do, and it is important enough to patients that we would say we don't need the evidence to move forward with it.

Any comments about that before we vote?

Stu?

MEMBER REYNOLDS: So, I would argue that, even though the data is not presented, that it is considered the standard of care to do cystoscopy at the time you do sling, and that certainly the risks of doing that and potentially the cost are far outweighed by the risk of not doing it. So, the benefits are far outweighed by the risk of not doing it.

CO-CHAIR SAIGAL: Thank you.

Personally, I agree. I think if my mom was having the surgery, for sure I would want her to have a cystoscopy. It is a low-risk thing to do, and it definitely catches things that are serious. So, my bias is that it is important, but the evidence is not super-strong.

Okay. So, should we vote?

MR. WILLIAMSON: Yes, we will now vote on the potential exception to empirical evidence. Again, this will be a hand vote.

For yes, please raise your hand.

(Show of hands.)

And for no?

(Show of hands.)

Okay. So, we have 13 yes and 2 no.

CO-CHAIR SAIGAL: Okay. So, then, the last part is the gap.

MEMBER GILL: So, in the gap data, they don't really present a gap, that people are doing it or not doing it. They just presented several articles that have suggested it may not be necessary. So, it is a little difficult to see the data on the gap performance to me, although it is implied that there might be one.

CO-CHAIR SAIGAL: Okay. So, it is, I guess, inconsistently presented in the document. I think my reading of the studies about whether cystoscopy is beneficial or not, some people are not doing it. So, that is just logical, I guess. But they do have the

observational data that like a third of people use routine cystoscopy. So, I think it is probably, my reading, there is a gap.

Any other comments?

(No response.)

Okay. So, let's vote as to whether there is a documented gap in the performance of this measure.

MR. WILLIAMSON: We will now vote on the performance gap. There are four options.

You may begin now.

(Vote taken.)

And we have 2 high, 7 moderate, 4 low, and 2 insufficient.

CO-CHAIR SAIGAL: Okay. Then, the next vote will be on whether the concept should be approved. So, to summarize it, it is looking at cystoscopy after the surgery. We think it is important to measure. We felt that the evidence supporting the measure was insufficient according to NQF standards, but

we felt that the problem was a big enough problem for the public that it was worth it to circumvent the NQF standard.

And we felt there was a gap. Generally speaking, the gap wasn't documented at a high level but at a moderate level.

And now, we can say if we think we should approve this and move forward to stage two.

MR. WILLIAMSON: We will now vote on the overall recommendation. This is a yes/no vote.

You may begin now.

(Vote taken.)

We have 11 yes and 4 no.

CO-CHAIR SAIGAL: Okay. That was the last -- are there any comments for the developers?

MEMBER PELLETIER-CAMERON: Just in their exclusions, in this one specifically, they state that concomitant prolapse surgery is an exclusion, but I don't see why that

would make you an exclusion, because what we are hoping to measure is that if there is an injury from your stress incontinence surgery, and I don't see how prolapse surgery factors in it.

I know it does in decisionmaking and treatment plans, but I don't see how having a prolapse surgery would make the use of cystoscopy any different. So, I disagree with that as an exclusion.

CO-CHAIR SAIGAL: Okay. So, the developers should just note that, once again, the exclusions in the document are being questioned.

Zahid?

MEMBER BUTT: I think it would be good to provide evidence for the gap because I think that is a gap.

(Laughter.)

CO-CHAIR SAIGAL: Okay.

MEMBER BUTT: And I voted no, to be consistent because I voted no on the gap

question.

CO-CHAIR SAIGAL: Okay. Got it. So, the document could be attended to, reorganized, and gap information could be strengthened.

Liliana?

MEMBER BORDEIANOU: So, the question I have is whether we should specify which of the two approaches. One is much more standard of care than the second one.

CO-CHAIR SAIGAL: For prolapse?

MEMBER BORDEIANOU: All right. So, do I have to wait for the green light to go on?

CO-CHAIR SAIGAL: No.

MEMBER BORDEIANOU: So, the only question I had is whether or not we should have a specification about which approach is being used, because it sounds like cystoscopy is much more important in one versus the other.

CO-CHAIR SAIGAL: Okay.

MEMBER BORDEIANOU: And that might be the issue about the gap and where it is standard of care and where it is not.

CO-CHAIR SAIGAL: Okay. So, a suggestion that the measure be divided specifically by procedure for the developer to consider as a plus or minus.

Ed?

MEMBER GILL: And then, also, for the developer, going forward -- this would be for stage two -- in terms of feasibility and usability, I don't put myself out there as a coding expert, but we talked about this a little bit before. I am not sure you are going to capture all the cystoscopies that were done if you use CPT codes because it may be bundled and not show up that it was done at all. You may not be able to measure it using your criteria.

CO-CHAIR SAIGAL: Okay. So, a caution on feasibility.

Okay. So, with that, then, there

is a member and public comment period. And so, I would invite anyone monitoring us --

MS. WILBON: Let's start with people in the room. If you have any comments or would like to address the Committee, please queue at the microphone. And then, if there is no one in the room, then we will go to the phone.

(No response.)

It looks like no one in the room.

Operator, Arnika, if there is anyone on the phone who would like to make a comment, could you please give them instructions on how to address the Committee?

THE OPERATOR: At this time, in order to ask a question, press *, then the number 1 on your telephone keypad. We will pause for just a moment to compile the Q&A roster.

(Pause.)

Again, to ask a question, press *, then the number 1 on your telephone keypad.

(No response.)

CO-CHAIR SAIGAL: Okay. So, hearing no comments, maybe we can have lunch. And so, 12:45 is lunch, is a 15-minute break for lunch. So, we have a working lunch?

(Whereupon, the foregoing matter went off the record for lunch at 1:16 p.m. and went back on the record at 1:36 p.m.)

A-F-T-E-R-N-O-O-N  S-E-S-S-I-O-N

1:36 p.m.

CO-CHAIR SAIGAL:  Do we have the AUGS representative here?  Great.

Okay.  So, our next set of measures --

MR. AMIN:  Chris, I think we still have one more AUA one.

CO-CHAIR SAIGAL:  Oh, I'm sorry. That is an AUA measure.

MR. AMIN:  2054.

CO-CHAIR SAIGAL:  Sorry about that.  I'm sorry.  We will call you up in a minute.  Okay?  We jumped the gun.

So, Stu, C2054, assessment of treatment within one year of SUI surgery.

MEMBER REYNOLDS:  You were just hoping that we had finished with all the stress incontinence ones.

(Laughter.)

So, this is Measure No. 2054, which is entitled, "The assessment of

treatment within one year of stress urinary incontinence surgery".

It is very similar to the ones that we have been discussing. Here, the numerator is the number of female patients who have had stress incontinence surgery and then received four components of their postoperative assessment within a year of their surgery: characterization of incontinence, physical examination, post-void residual analysis, and urinalysis urinary culture. This is very similar to the ones that we had discussed previously in the preop assessment.

The denominator statement is those who underwent stress incontinence, and there are, again, very similar exclusions to the ones that we had addressed before.

Again, this is a process concept. I think it is important because, in terms of the outcome, it is, again, not entirely clear to me what the overall outcome is or exactly

how the process relates to that, but we will probably discuss that a little bit more as we go forth.

I would draw your attention, there were a number of comments made by the technical staff regarding this one, as well as comment from the members. And then, there were some previous ratings and comments. They are on some of those forms. I won't read them for you.

But I guess I would say that our first task is to look at the impact. And again, similar to the previous ones, there are a number of studies that are included that suggest that stress incontinence is common and, thus, has a great impact on not only U.S., but also worldwide populations, although the evidence does not specifically relate to the measure, but mostly to the idea of stress incontinence, and that stress incontinence surgeries are commonly done.

CO-CHAIR SAIGAL: So, Stu, besides

the evidence, just we are going to be voting on the impact, I suppose.

MEMBER REYNOLDS: Yes.

CO-CHAIR SAIGAL: So, could you just summarize, is this an important national health goal? Does not have a lot of people that are affected by it? A lot of resources that are used to treat it?

MEMBER REYNOLDS: Yes to all the above questions.

CO-CHAIR SAIGAL: Okay.

MEMBER REYNOLDS: And they give specific details which are, again, the same ones that have been on the other ones. But anywhere from 4 to 30 percent of the female population, direct costs of over $13 billion back in 1995, and it has, presumably, increased since then. So, yes, I would say at least certainly moderate to high impact --

CO-CHAIR SAIGAL: Okay.

MEMBER REYNOLDS: -- would be my recommendation.

CO-CHAIR SAIGAL: Andy?

CO-CHAIR BASKIN: Well, I would argue that this measure, the impacts already occurred before this measure. I mean, the costs are all gone. I mean, that has all been done. Now we are measuring whether you should be doing something a year later. I am not so sure what the high impact is of people who have already had surgery. That impact is not part of this, you know, a year later getting an assessment. I don't necessarily read that as being high impact.

CO-CHAIR SAIGAL: I think if you look at the way it says here, "a specific national health goal or priority," so I think measuring outcomes of care is a health priority for us. Would you agree? And certainly, treatment efficacy is part of that. So, in my mind at least, having attention to outcomes is a health priority.

DR. PACE: Except this measure is not about outcomes. It is about another

process of assessment.

CO-CHAIR BASKIN: All right. What is about the characterization of incontinence, a history, of physical, a post-void analysis, that is truly a measure of outcome? I mean, this is the real outcome, is patient satisfaction, patient functional status. I mean, you know, stuff that is beyond what they are asking here.

CO-CHAIR SAIGAL: Well, they are characterizing incontinence in this measure postoperatively. I mean, that is the medical outcome of the measuring.

DR. PACE: Right. So, for us to consider this an outcome measure, it would be something like percent of women who have the surgery who have their incontinence resolved. That would be the outcome measure.

So, this is really about the process of doing an assessment. One of the rationales given was so that you can have some outcome data, but that is not an outcome

performance measure. It is still just the process.

CO-CHAIR SAIGAL: Okay. So, to be clear, this is still a process-of-care measure that relates to a potential outcome-of-care measure which would be measuring the success of the surgery. So, in my view -- and anyone who wants to, please jump in -- the way that it meets high impact is because it is a process that allows us to measure outcome, which is a health goal and priority for the country, in my view.

But Andy and Karen feel that it is not necessarily a process that matters because the surgery has already happened, and that we are not doing something to actually say, was the surgery success or not?

Andy, am I getting you right?

CO-CHAIR BASKIN: No, I agree. I think an assessment a year later is a good idea, but if the assessment is truly a measure of success, not a measure of documentation

that you ask the question.

CO-CHAIR SAIGAL: As written, it is a process measure. You don't have to say whether the person was continent or not.

Okay. And are there other comments about the impact or the nature of this kind of -- Judith?

MEMBER TOBIN: Because I am trying to evaluate the process of evaluating these measures, isn't part of evaluating a process measure its proximity to the outcome? So, then, the issue that Andy spoke to earlier would be important if you don't think that proximity, you know, if it is too far away from the outcome to have any kind of an impact. That was something, as I went through these, I was thinking about as well.

CO-CHAIR SAIGAL: I think that Andy was talking about the surgery already happened. I think what you are talking about is it is a process-of-care measure that is, I think it is proximate to measuring an outcome.

I mean, it is within a year. If you ask a person, "Are you still incontinent," and they say yes, you can then make the link and say, well, you are going to record that in the database and have an outcome. So, I think that is a pretty proximate measure.

But what Andy was talking about is, is it worth it to measure a process-of-care measure after the surgery is done? I think that is the distinction that we have to decide on, where we lie.

Okay. Yes, Paul?

MEMBER MERGUERIAN: If you look at the numerators, they are pretty much the same measures that they had in 2049, which is a preoperative assessment, except for one measure, basically. So, it is really looking at the characterization of the incontinence. If they are incontinent, it is a focused physical, post-void residuals, and urinalysis.

CO-CHAIR SAIGAL: Yes, right. And what is your conclusion from that?

MEMBER MERGUERIAN: Well, it should be part of the examination. The question is if we link that to an outcome. I mean, I have a hard time trying to figure out those process measures without linking it to an outcome, because we have got to link process with outcome. And what is the outcome that they are looking it? Is it really success of surgery?

CO-CHAIR SAIGAL: Yes, that is a good question. That is basically the idea. Is this process-of-care measure important to measure because the implication is you would learn if the person was incontinent or not after surgery. So, if you believe that that is a reasonable thing to do as a measure, then you vote this as high impact. If you think it is not important enough to do that, you should measure directly the outcome, then you can vote a different way.

MS. WILBON: I just wanted to point out that there is a question on the form

that asks the developer to demonstrate where they feel like this measure falls in that stream, that value stream. It is 1(c)(3). So, if you are trying to figure out which outcome they felt like they were measuring to be proximal to, it will give you an idea of what their value stream was for mapping the process/outcome linkage.

CO-CHAIR SAIGAL: Uh-hum. Okay. Yes?

MEMBER BORDEIANOU: What I thought I heard is that the measure also looks at complications of surgery. So, measuring the post-void residual shows you whether or not the sling is too tight. And so, it is a process of quantifying complications.

CO-CHAIR SAIGAL: It is a process that you would need to do to get an outcome like a complication rate.

She is saying that PVR, measuring the post-void residual, if you have a high post-void residual, then the person is in

retention, and that is possibly as a result of the surgery causing the sling to be too tight. So, again, it is a process that measures a potential outcome.

Robert?

MEMBER ELLIS: Since this is being measured using administrative data and paper records, is there a specific identifier for this type of followup that would identify specifically what we are trying to measure here?

CO-CHAIR SAIGAL: Well, I am not sure that we are supposed to get into feasibility of it.

MEMBER ELLIS: Okay. I'm sorry.

CO-CHAIR SAIGAL: But I think that, just in general, some of these things are capturable by CPT codes, yes, PVRs, physical exams, urinalysis and culture. Obviously, the characterization of incontinence is not. It is from a CPT II code.

DR. PACE: In general, the CPT II codes are the physician saying that they did these four or five. So, it would be giving a code that says I did this post-surgery assessment, is generally the way they are structured.

But, in this case, again, this is a concept. So, we don't have that detail yet of what the CPT II code would actually be.

CO-CHAIR SAIGAL: Okay. Can we vote then whether this is high impact? Can we vote as to whether this is high impact or not?

MR. WILLIAMSON: We will now vote on the impact. There are four options: high, moderate, low, or insufficient.

You may begin voting now.

(Vote taken.)

I think that is every single one. Zahid is out of the room. Okay.

And we have 2 high, 7 moderate, 4 low, and 1 insufficient.

CO-CHAIR SAIGAL: Okay. So, then,

I think it proceeds. So, now you talk about the evidence underlying the measure.

MEMBER REYNOLDS: So, in the portion of the document in which there is and should be evidence, there is essentially no evidence listed. That would be in the pilot submission form under Sections 1(c).4 all the way to 1(c).13.

I don't have a great opportunity to explain this. There is some data discussed in other sections. This is a problem we have run into before.

I am not aware specifically of data that suggests, if you do or do not check, it would affect the outcome of the measure. I think that is either difficult to measure or maybe hasn't well been done. Certainly, the AUA guidelines, which have been previously referenced and are referenced, recommend for efficacy studies that you should have at least 12 months followup data for that. But I am not sure there is any other data that suggests

there is a specific --

CO-CHAIR SAIGAL: Okay. So, this is one of the situations, I think, unless there is other data that people are aware of that is relevant here, where this is essentially a consensus statement from experts that you should measure or you should ask patients about how they are doing after surgery and examine them. So, I think probably we would be in the situation we had in the past in this vote.

Are there any other studies people are aware of that are relevant regarding physical exam after surgery?

(No response.)

Okay. Then, I would suggest we move to a vote on the evidence, whether, yes, it does meet the NQF criteria; 2, it does not, or 3, there is evidence around us that we are aware of that they are not aware of.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

CO-CHAIR SAIGAL: Okay. So, it doesn't meet the NQF criteria.

MR. WILLIAMSON: Just for the record, we have 13, no, that the evidence does not meet the guidance for the quality, quantity, consistency, and 1, no, that there is insufficient information submitted.

CO-CHAIR SAIGAL: Okay. The person who voted insufficient, is that because you are aware of other data? Or was that just a misfire? Or don't want to admit it?

(Laughter.)

Okay. So, then, in terms of the next vote, it has got to be as to whether we think that, if a person has had surgery, that quantifying their incontinence after surgery and doing a physical exam is important enough and has enough face validity that we would skip evidence to support it.

And again, we are not talking

about getting an outcome recorded, but just doing a process to collect data. That is what we are measuring. So, that is the question for the group.

Any comments about that? Yes?

MEMBER BORDEIANOU: I would say that, if we are reporting the data, we would want to report much more data about an outcome with surgery than just that, complications, et cetera.

CO-CHAIR SAIGAL: So, you conceptually could have more data?

MEMBER BORDEIANOU: Incontinence. Yes, more.

CO-CHAIR SAIGAL: Okay.

MEMBER BORDEIANOU: I don't think it is sufficient.

CO-CHAIR SAIGAL: Stu?

MEMBER REYNOLDS: I have a little bit of issue in sort of how they are defining the timing of it. And I don't know if that comes up now or at another time. I mean, it

is within 12 months. Admittedly, I think it is hard to know what the correct timing would be, but there is a big range between one day and 12 months, if you are looking at outcomes. And so, I think there is a flaw here in terms of the timing of that. I don't know what the right answer is, but I see that as an issue.

CO-CHAIR SAIGAL: Okay. So, I think these two comments have to do with our general sense of how important this measure is. If it was a great measure without a lot of concerns, maybe we skip over the fact that there is no evidence to support its implementation. But if there are concerns about the measure itself in terms of how it is constructed or the value of the data it provides, that is something we have to consider as well.

Any other comments about this measure?

(No response.)

I will say that my take on it is

that it is important to measure. I think that it would be nice to link it to some harder recording of what the result was about the patient was incontinent or not incontinent in the chart. But that is just my view.

All right. So, we should vote. So, this is a hand vote?

MR. WILLIAMSON: I think it is fixed now.

CO-CHAIR SAIGAL: Oh, good. Okay.

MR. WILLIAMSON: We will try this, and if it doesn't work, then we will do a hand vote.

But we will now vote on the exception to the evidence.

You may begin voting now.

(Vote taken.)

We stand at 14, yes. If everybody wants to point at the receiver again? There we go.

And we have 4 yes and 10 no.

CO-CHAIR SAIGAL: Okay. Then, it

does not pass the exception measure. So, then, it doesn't proceed. Okay.

And then, are there comments for the developers that we ought to give them?

So, I could summarize what I have heard. Mainly, the concern was that just having a process-of-care measure at this point in the evaluation was a little weak, and that having an outcome measure would be better; that to look at outcomes, maybe there are more data points that you would want to collect specifically from patients in terms of infections, or whatever it might be, and that the window of assessment could be better defined. It is a little broad from day one to day 365.

Anything else? Comments for the developer?

(No response.)

Okay. Good.

Then, the next one -- yes?

MEMBER SCHOENFELD: Evan, could

you go back one slide? Right. No. Right there.

I just wanted to reaffirm something that I think I heard from Karen and Taroon before, even though I got here a little bit late. We have already come to this kind of a vote a few times. And I just wanted to make sure I understood.

Voting yes to this kind of question should truly be unusual and uncommon. Is that the overarching philosophy about this? I mean, I am a gastroenterologist, not a GU specialist. I am not trying to step on anybody's toes. But I just want to make sure I am reaffirming what I heard was the philosophy with respect to this situation.

MR. AMIN: We will tag-team here.

I think, clearly, yes, it should be an exception and there should be a compelling reason why you should vote yes.

And there have been discussions earlier today in which there have been, in the

sense that a randomized controlled trial or any other experiment would be clinically unethical or that the benefits clearly outweigh the harms in terms of the benefit to the patient.

So, really, this is intended to be an exception. I mean, the challenge with invoking this frequently is that you can make an argument that a measure that is really distal to an outcome would not have evidence as well. But what we are really trying to do is get measures that are closer or more proximal to outcomes.

So, when you invoke this, it becomes very unclear to those external on which one it is. Is it just so distal that it won't have evidence or that there isn't evidence because of some exceptional reason, but the benefits greatly outweigh the harms.

And so, there needs to be an explanation, a clear rationale from the Committee that, when you are invoking this,

why, very clearly why, and what the benefit to the patient is. And there should be a high bar for why you let measures go forward under this exception.

But, Karen, if there is anything --

MEMBER SCHOENFELD: When I heard you say that, it just kind of reminded me to expand on what John said. I mean, yes, we don't need a randomized controlled trial about the efficacy of parachutes when you jump out of a plane. On the other hand, I am not sure we need a quality indicator to say, yes, you should wear a parachute when you jump out of a plane.

DR. PACE: And I think that is the crux of it. First of all, I will keep emphasizing our criteria do not require randomized controlled trials. We recognize all types of evidence. But I think the real issue is trying to, as Judy mentioned, our guidance is to really focus National

Performance Standards on those things that are most proximal to the desired outcome. Because, as you know, once you assess something, somebody has to actually interpret that assessment. They have to identify potential treatment alternatives. They have to discuss those. The right treatment has to be applied before you really get to the outcome.

And so, measuring something way down here does not in any way assure that we are going to get to the outcomes. And so, the push from NQF is for National Performance Standards to measure those things that are most proximal. It doesn't in any way say that assessment isn't important; it is critical. You all have to do it in everything that you do. But we are trying to get at those things that are most important in terms of National Performance Standards. Because, as you can all imagine, there are thousands of things that you do that could be measured if we start

talking about all the different assessments that all the different patients could have. And that is really the emphasis.

CO-CHAIR SAIGAL: Okay. So, then, that is that for that measure.

Okay. The next one, Stu, it is yours still, C2037.

MR. AMIN: But you want the --

CO-CHAIR SAIGAL: Oh, yes, yes, yes.

AUGS, could you please come up and introduce your measures?

MR. AMIN: This is for the set of measures. I just want to make that clear.

CO-CHAIR SAIGAL: The next three measures, two or three minutes of your time, please. Thanks.

MS. PULLIAM: My name is Samantha Pulliam. I am from Massachusetts General Hospital, and I am here to present the next three measures for the American Urogynecologic Society.

The first measure is evaluating the percentage of female patients with a characterization of the degree or prolapse in each vaginal compartment using a validated, objective measurement system. And we would like that to happen within 12 months of surgery.

Annually, there are over 200,000 surgeries to repair pelvic organ prolapse, and there is a great diversity of surgeries addressing a variety of problems within the pelvis. Failure to fully evaluate prolapse causes inadequate or incorrect surgical repair, resulting in increased patient morbidity, failure of the surgeries, and reoperation rates that are increased, and often resulting in unneeded expense.

The POP-Q, which stands for Pelvic Organ Prolapse Quantitative System, and then the Baden/Walker, which is the antecedent to that, are validated, objective ways to evaluate all areas of prolapse in the vagina.

These measurement systems are taught in residency. But only 43 to 78 percent of people who could use these tools do so.

The evaluation systems are endorsed by the American Urogynecologic Society, the International Urogynecologic Association, the International Continence Society, and the NIH. We believe that universal use optimizes communication and treatment for these women.

The POP-Q system, for example, is an evaluation using nine points within the pelvis to quantify the anterior vaginal wall, the posterior vaginal wall, and the top of the vagina and the uterus.

This is a high-impact measure in that evaluates properly over 200,000 women who are going to undergo surgery. There is a gap in that maybe up to 50 percent of specialists don't use this. And this is a validated tool that is recommended by multiple societies.

The second measure looks at

suspension of the top of the vagina during surgery for pelvic organ prolapse. To perform a hysterectomy alone on a patient who has pelvic organ prolapse is to leave the apex of the vagina unsuspended and does not repair the problem.

Over 78,000 hysterectomies are performed each year for pelvic organ prolapse, and perhaps only 53 percent of patients have vaginal apical suspensions at the time of this hysterectomy. This means that there is an increased risk of repeat surgery; 7.4 percent of those with hysterectomy alone as compared to 2 percent who have a complete prolapse repair require repeat surgery.

This is an American College of Obstetrics and Gynecology guideline. It is interesting in that there are multiple systematic reviews looking at which type of apical suspension is appropriate. And perhaps most impressive for its absence, hysterectomy alone is not a type of suspension for the

vagina apex. It would be unethical, we think, to perform a randomized trial evaluating the absence of an apical repair to address prolapse.

Again, a high-impact evaluation. The gap is present and has been demonstrated, and the evidence has been evaluated in multiple systematic reviews.

The final measure is looking at patients who undergo cystoscopy at the time of surgery for correction of anterior or apical vaginal prolapse. Damage to the ureters and the bladder at the time of prolapse repair surgery can occur at a rate of 5.1 percent. In many studies, only as few as 12 percent of ureteral and 35 percent of bladder injuries are detected without cystoscopy. And we know that unrecognized damage is dangerous and expensive and may result in hospital readmissions, repeat surgeries, and even renal demise.

Routine cystoscopy identifies 99.4

percent of bladder of ureteral injuries, and is cost-effective for an injury rate of 1.5 to 2 percent. This is a guideline that is endorsed by the American College of Obstetric and Gynecology. This is a procedure for which ACGME residency programs in both obstetrics and gynecology provide training. In most operating rooms, cystoscopy equipment is readily available.

So, again, this is a high-impact procedure. There is a defined gap of people who fail to perform this procedure, and there are systematic reviews that contain strong evidence supporting routine cystoscopy for procedures repairing prolapse.

CO-CHAIR BASKIN: Thank you.

Stuart? Oh, okay. So, Stuart, you are the lead discussant here?

MEMBER REYNOLDS: Yes.

So, this is specific to the first concept, which is No. C2037, objective characterization of pelvic organ prolapse

prior to surgery. We had a good overview there.

I would point out that the numerator for this concept is the number of patients whose pelvic organ prolapse exam was documenting using a validated tool as described, either the POP-Q or the Baden/Walker, within 12 months prior to surgery for pelvic organ prolapse. And then, the denominator is all the patients or women who have undergone pelvic organ prolapse surgery.

Again, this is a process concept. There is a component of the relationship between this process to an outcome measure, and it is suggested that proper documentation of the degree of prolapse will likely increase the chance of the appropriate procedure and, thus, reduce the risk of recurrent prolapse later.

I guess getting to the meat of what we are here to discuss is the impact.

Pelvic organ prolapse is a common condition. There are a lot of surgeries that are being performed in the U.S. for it. There is some increased impact or awareness because of ongoing issues brought on by the FDA over the last couple of years regarding the common materials used for these types of repairs, specifically the synthetic mesh that we have been talking about before, but plays a bigger role in the pelvic organ prolapse.

And so, the impact, for example, of getting the diagnosis correct maybe is even more important here because the negative aspects also can be so high. And so, I guess looking at the evidence for high impact, well, I guess we could open that up for discussion, but it looks like there is probably something in terms of moderate, in terms of how much prolapse plays a role.

It is not clear from the evidence, the summary of the evidence for high impact exactly what the role of the POP-Q or the

Baden/Walker assessment is in that, but that is, I presume, a presumptive relationship.

CO-CHAIR SAIGAL: So, thank you.

Any comments here? I mean, it sounds like there are two parts of this. One is that this characterization occur. And two is, should it occur with one of these objective tools? I don't know whether people want to speak either one of those. I am not an expert here, other than, once again, this is one of those did you take a history; did you characterize it happening. I can't speak to the value of these tools versus an alternative. So, if someone could, that would be very helpful for me.

MEMBER REYNOLDS: Well, I would say that they are the most common methods or most common validated methods to assess that. So, generally accepted, the POP-Q is the most common and the Baden/Walker. So, it is the grading system that you would use for a prolapse.

MEMBER BORDEIANOU: This really gets to the crux of who needs surgery and who doesn't, because there are lots of women with prolapse that don't necessarily need surgery as the first step. By quantifying prolapse introitus, either with a Baden/Walker or with a POP-Q, you stratify that better.

CO-CHAIR BASKIN: Yes, but, once again, we get into the chicken-and-the-egg thing because this measure is a measure of those that have had surgery, not as the measure of whether someone should have surgery or not. So, it is similar to the other issue we had. It is having occurred prior to surgery, but the denominator is only people who have had surgery. So, the construct may not be getting the best value for this.

MEMBER REYNOLDS: It is a different measure if it is all women with prolapse, and then you could discuss their treatment options. But you're right, this is specific for surgery.

MEMBER GILL: So, I think one of the differences with this measure and the other one that we were discussing is this one has clearly been shown -- and I know we are not on gap and everything -- but it has been clearly shown to make a difference. If you don't do this preoperatively, get the right diagnosis, then your outcome is going to suffer. So, I mean, I think this is different than the other one because there actually is evidence to support this.

CO-CHAIR BASKIN: Is it the right diagnosis or the severity of the diagnosis that this tool is doing? I am really asking that. Because if you think someone has prolapse, I am not sure how that diagnosis is made. And the tool is really to assess the degree of prolapse that would help you, then, choose? Is that what happens here?

MEMBER GILL: Well, I think there are a couple of things here. There is grading it and recognizing. I think that is part of

the measure, is that, if it is not measured, it may not be recognized and, therefore, it wouldn't be treated.

MEMBER REYNOLDS: Both grade and anatomical defect. So, if you asked a woman if she has a bulge, it is very difficult to delineate what that is that is prolapsing in. If you just looked without doing a more standardized exam, you would just see something bulging out, but, again, you may not know if that is anterior defect, an apical defect, or a posterior defect, which is how they describe how you can divide the vagina into compartments. And all three of those different anatomic defects, which may present the same way, have different treatment options. So, knowing exactly what you have would determine the type of, for example, surgery that you would perform.

CO-CHAIR BASKIN: Any other comments?

(No response.)

Then, I think we will go ahead to vote on this one.

It sounds like this is one where the actual assessment could make a difference in the choice of treatment. I think that is, to me, where the quality part comes up, but I guess we are not up to that part yet, to the evidence part. So, this is just the impact part. So, I'm sorry, I jumped ahead.

(Laughter.)

MR. WILLIAMSON: We will now vote on the impact. There are four options: high, moderate, low, or insufficient.

And you may begin voting now.

(Vote taken.)

And we have 9 high, 5 moderate, zero low, and zero insufficient.

CO-CHAIR BASKIN: Okay. Thank you.

And then, we will move on to the quantity and quality of the evidence.

Stuart, did you want to make

another comment regarding that?

MEMBER REYNOLDS: So, in terms of the evidence, there is quite a bit of evidence that is listed for us to review, many of which are of pretty high quality. The issue I have with the evidence is that I am not sure it specifically addresses the measure as it is described. So that, specifically, the evidence that is presented is about the study, the diagnostic procedure itself. So, how good is, for example, the POP-Q at detecting the grade and anatomical effects?

And then, there are a lot of studies looking at correlation between observers and whether it is repeatable. And all that is high-quality evidence, and good data suggests that the test is reproducible and that seems to measure what it does.

I don't see that there is any evidence here suggesting that, again, by assessing that, you are going to change the downstream outcomes. This is the issue that

we have been running into all day. But there is a lot of data here, but, again, I am not quite sure it is exactly applicable to the measure.

CO-CHAIR BASKIN: So, it sounds like the data is validation of the tool?

MEMBER REYNOLDS: Correct.

CO-CHAIR BASKIN: Okay.

CO-CHAIR SAIGAL: But not validation of an outcome in matters of patients.

MEMBER REYNOLDS: Yes, I agree.

CO-CHAIR SAIGAL: Just a similar issue.

MEMBER REYNOLDS: We are in the same boat.

CO-CHAIR SAIGAL: Yes.

CO-CHAIR BASKIN: We will do this vote first, and then, depending on the results of this, we have an optional vote that comes up next.

So, I think we are ready to vote

on this one. I think the statements have been made that the evidence really is not here for the actual outcome, but just that the tool does what it is supposed to do, measures what it is supposed to measure. So, vote with your heart -- and the facts.

(Laughter.)

MEMBER BORDEIANOU: What is the outcome?

CO-CHAIR BASKIN: I guess the issue is, does doing this have any effect on the type of surgery or lack of surgery or what kind of surgery is performed, which I think was really what the goal is, is that you are going to get a better outcome because you have done the right surgery.

They talk about reoperations is the issue here. And I don't see that the evidence does anything for that.

MEMBER BORDEIANOU: That is just saying doing a good physical exam, when calling it a POP-Q or Baden/Walker, and using

a language to describe your exam. I mean, do we need data to document the physical exam?

It is like the discussion we had about urinary incontinence. It goes back to the same thing. There might not be a study, but it is intuitive that doing a good physical exam before doing a surgery and planning a surgery appropriately would improve outcomes.

CO-CHAIR SAIGAL: Yes, I mean, the idea is that you are believing that doing the systematized exam will have an impact that is beneficial for the patient, which I think sounds obvious, as you are saying. But there is no evidence that they have presented to support it. So, we are in that situation of just believing the expert consensus, that if you do this, they will get better care.

CO-CHAIR BASKIN: So, that would be the next vote, if we get that --

MEMBER MORTON: I was just going to mention having an H&P as part of the record is a CMS requirement. So, this goes a bit

beyond just an H&P.

CO-CHAIR BASKIN: So, let's vote on this part of it, which is clearly the evidence that is submitted and is it evidence for the health outcome or not. And that is what this vote is about. So, that is going to be a 1, 2, or a 3? Is that it? Can you put that back up there for us? And let's bring this to a vote. I think we all know what 3 is at this point in time, too.

But I don't think anyone has submitted any evidence to say that it is out there, that wasn't reported to us. So, I don't think that is a great option for us.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

CO-CHAIR BASKIN: Okay, that is pretty clear.

MR. WILLIAMSON: We have 1 yes; 12, no, that the evidence does not meet the

guidance, and 1 that, no, it was insufficient information submitted.

CO-CHAIR BASKIN: So, now we have the branching question comes out, that only if we say no here can we get this question. I think that is what we all really wanted.

And that is whether there is a compelling exception here. I am going to ask people if they want to speak to this. I understand we have made an argument before about this, but I think people ought to realize, though, that there is an opportunity to have a better measure that actually does measure outcomes, as opposed to whether you just use the tool or not. And no one has taken the opportunity to create that measure. But is this the low-bar measure that we should accept or is this the measure that is the better measure that could have been --

CO-CHAIR SAIGAL: So, what you are saying is that you would prefer an outcome measure of appropriateness of the surgery?

CO-CHAIR BASKIN: Well, that would seem more proximal to what we wanted, yes.

MEMBER PELLETIER-CAMERON: And pelvic organ prolapse is a little bit more complicated than that. Just because you have a prolapse doesn't mean it needs to be repaired. The POP-Q does not evaluate "bother". I mean, for most people who do this surgery, it really doesn't matter to me what their POP-Q score is. I mean, that does matter to me and I evaluate it, but what makes me decide whether or not to do surgery is the "bother". I do a POP-Q on every patient that I operate on because I think that it is a great measurement tool, but I don't think you could use the POP-Q to decide whether or not surgery was appropriate.

CO-CHAIR BASKIN: And that is not what the measure is.

MEMBER PELLETIER-CAMERON: Right.

CO-CHAIR BASKIN: It is these people have had surgery. That is why they are

in the denominator of the measure.

So, my point is, why not measure whether they had the appropriate surgery as opposed to whether they just had surgery? Isn't that what the tool is supposed to help decide, is what is the appropriate surgery?

MEMBER FALLER: To go back to what Anne said, it is not just did they have the appropriate surgery based on their POP-Q, or whatever you are using, but did they get the solution that they wanted? I mean, they were bothered by it. Are they no longer bothered by it? It is like you were incontinent; are you no longer incontinent? You were bothered by it; are you no longer bothered?

MEMBER GILL: So, the other thing that this tool is useful for is communicating between physicians, evaluating further studies. So, it is not just a patient outcome and question for that individual patient, but sort of on a more global scale, to see if these surgeries are going to be appropriate.

I am not sure how this fits in here, but it is certainly an important thing to do on a larger scale. It is a measure of quality to see if that was done or not and to be able to evaluate things more globally later on. So, again, I am not sure how it fits in, but that is one of the main ways we use this.

DR. PACE: Because I know that is part of the rationale that was laid out, is that the reason for this is to have data for something later, you need to evaluate that in terms of we endorse performance measures for accountability based on things that really should be done. So, we don't endorse measures for research, is what I am saying. You know, it is really for accountability applications and performance improvement. And obviously, there are linkages there.

MEMBER BORDEIANOU: I only wanted to say that, also, the "bother" can be a wrong thing to measure as well because there are a lot of patients with pelvic organ prolapse

that have a lot of other psychiatric issues and emotional issues and associated issues. You need some quantifying measurement of the disease, whether it exists or it doesn't, whether there is a component of constipation that is going on here as opposed to prolapse, et cetera.

CO-CHAIR BASKIN: Okay. Well, if there are no more comments then, I don't know that I can summarize that as pushing us in one direction or the other here.

It is not the most proximal outcome. But the question is, is it valuable to measure this and is there a compelling reason to think that this will improve outcomes, in and of itself?

So, let's bring that to a vote.

MR. WILLIAMSON: All right. We will now vote on the exception to empirical evidence. The question posed is, is there an exceptional and compelling reason that the measure should be considered further?

And begin voting now.

(Vote taken.)

We are missing two, if everybody wants to point at the receiver again and vote. There we go.

And we have 3 yes and 11 no.

CO-CHAIR BASKIN: All right. Then, I guess we move on.

CO-CHAIR SAIGAL: Any comments for the developers?

CO-CHAIR BASKIN: Oh, yes. I'm sorry. Please. Anyone want to make comments to the developers, I mean other than the obvious, what we have already stated in that there is a more proximal outcome here which may be a more meaningful measure?

(No response.)

No other comments? Thank you. Then, we will move on.

CO-CHAIR SAIGAL: So, the next one is C2038, performing vaginal apical suspension at the time of hysterectomy to address

prolapse.

And Dr. Gill has that for us.

MEMBER GILL: So, the crux of matter here is patients that have uterovaginal pelvic organ prolapse that are undergoing surgery for it. They are broken down into two main categories, the women who just had hysterectomy as the treatment and women who had hysterectomy plus a specific additional procedure to support the apex.

It is an incredibly-common procedure, you know, 100 to 200 thousand done for prolapse a year. It is very costly. It affects a lot of people.

And we will eventually get to the evidence and the gap, but there is a lot of room for improvement here. I guess we will just talk, again, about impacts.

I think it is very high-impact. A lot of people, a lot of cases, a lot of surgery, and it is very expensive if we get it wrong.

CO-CHAIR SAIGAL: Okay. So, any other comments about the impact here? It sounds like it is a pretty important -- it is a very common women's health issue, and doing the right surgery is important to women.

So, can we vote on impact?

MR. WILLIAMSON: We will now vote on impact.

You may begin voting now.

(Vote taken.)

We have 13 high, 1 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. So, it passes that criteria.

So, Dr. Gill, could you tell us about the evidence that supports the measure?

MEMBER GILL: So, this actually has a lot of good evidence, including guidelines I will talk about first. ACOG recommends it. In addition to that, there are published systematic reviews, including randomized controlled trials. In addition,

there are five new reviews that have come out, all supporting this measure. So, to cut to the chase, there is a lot of evidence this one, actually.

CO-CHAIR SAIGAL: And the quality of the evidence, it is randomized and some observational?

MEMBER GILL: It would be categorized, I would think, moderate to high.

CO-CHAIR SAIGAL: Moderate to high? And a high amount of data --

MEMBER GILL: Uh-hum.

CO-CHAIR SAIGAL: -- all pointing the same way?

MEMBER GILL: I think so.

CO-CHAIR SAIGAL: And that is looking at the actual process. The outcome being -- was it failure of the repair?

MEMBER GILL: It is, for example, as was mentioned earlier, if just the hysterectomy is done, 7 percent of patients need a reoperation. If you do it with a

specific colpopexy procedure, only 2 percent need reoperation.

CO-CHAIR SAIGAL: Okay. So, that is failure. There is a direct link to an outcome that we care about?

MEMBER GILL: Right. Correct.

CO-CHAIR SAIGAL: Good. That makes it easy.

(Laughter.)

MEMBER GILL: Right.

CO-CHAIR SAIGAL: So, let's take a vote, then, on the evidence.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

And we have 14 yes.

CO-CHAIR SAIGAL: Awesome.

Okay, the next one is the gap.

MEMBER GILL: So, fortunately, we actually do have information on the gap as well that identifies a large gap. Based on

these recommendations, these guidelines, in some series, only 35 percent of surgeons are following the recommendations currently. So, there is a tremendous amount of room for improvement.

CO-CHAIR SAIGAL: Great. So, strong observational data from the California Hospital Survey about the gap and some other articles as well.

Any other comments about the gap?

CO-CHAIR BASKIN: A question about the gap.

CO-CHAIR SAIGAL: Okay.

CO-CHAIR BASKIN: So, is it because this is a newer guideline or a newer standard to do this, that it hasn't been widely adopted? Is that the reason? I mean, I can't understand why something that so clearly should be done isn't done all the time. I just don't understand it. Is it just something that, obviously, hasn't made its way through the community yet?

MEMBER GILL: Yes, yes. You know, I think it is a work-in-progress. Twenty or 30 years ago, hysterectomy was pretty much the accepted standard treatment for it. But, as more information has been garnered, it has changed. So, I think it has just been slow to be adopted, but it is clearly evidence-based.

MEMBER PELLETIER-CAMERON: And it is harder to do.

MEMBER GILL: That's true.

MEMBER PELLETIER-CAMERON: There is more skill involved. Doing a hysterectomy or doing a hysterectomy plus a proper apical suspension, it is just harder to do for the surgeons.

MEMBER BORDEIANOU: Does that mean that converts all the surgery to transabdominal versus -- no? Okay.

CO-CHAIR SAIGAL: Can do it robotically, in fact.

So, anyway, I think that we can vote then on evidence about the importance of

the gap.

MR. WILLIAMSON: We will now vote on the performance gap. And there are four options: high, moderate, low, or insufficient.

You may begin voting now.

(Vote taken.)

We have 12 high, 1 moderate, zero low, and 1 insufficient.

CO-CHAIR SAIGAL: Okay. Great. So, then, the general vote about approval of the concept.

To summarize, this is a process-of-care measure about a very important women's health problem. There is high-quality evidence or moderate-quality evidence that links it to an important outcome that patients care about, which is reoperation or treatment failure. And there is evidence that there is a significant gap in performance.

So, let's vote.

MR. WILLIAMSON: We will now vote

on the overall approval of the concept. This is a yes/no question.

You may begin voting now.

(Vote taken.)

And we have 14 yes and zero no.

CO-CHAIR SAIGAL: Okay. Any comments for the developer?

(No response.)

I congratulate you for a strong measure.

And then, the last one is C2063, and that is appropriate use of cystoscopy in pelvic prolapse repair, probably similar to the one we saw before.

Anne, could you talk about the importance to measure on that one?

MEMBER PELLETIER-CAMERON: So, similar to the previous discussion where we had whether or not you should do a cystoscopy at the same time as another surgery, whereas, here we are talking about doing a cystoscopy at the time of a prolapse repair. They are

specifically discussing cystoscopy at the time of an anterior or an apical suspension. And that is because, if you are just doing a posterior repair, there is very little risk of bladder injury. So, they are really focusing on the procedures that do carry a high risk of bladder and ureter injury.

To contrast that with the sling that we were discussing, there is really not a risk of ureter injury. But with any of these, especially the apex and the cystocele repairs, there is a risk of ureter injury.

So, their numerator is they are looking at the number of patients who have a cystoscopy at the same time as their apical or anterior repair, and the denominator is the number of patients who are having the repair. And they are not excluding anybody. They are not excluding sling patients. They are just saying, if you are having one of these repairs, you should have a cystoscopy done.

And this is a process. They are

identifying these two procedures by CPT codes.

I guess the importance of this is pretty clearly stated. There are several studies. I think I counted eight or nine studies, some observational, but all had fairly substantial numbers of patients stating that there is a real risk of injuring the bladder and the ureter during these procedures. And obviously, if you injure a ureter and you don't recognize it, someone could lose their kidney. And the same goes with the bladder; if you injure a bladder, you could end up with a fistula or other problems. So, I think the high impact of this problem was very clearly stated in the literature.

CO-CHAIR SAIGAL: Okay. So, in terms of any other comments about the importance to measure this, its impact on the population health, as a general concept?

(No response.)

Okay. So, let's vote on that then.

MR. WILLIAMSON: We will now vote on the impact. You have four options: high, moderate, low, or insufficient.

Begin voting now.

(Vote taken.)

And we have 8 high, 6 moderate, zero low, and zero insufficient.

CO-CHAIR SAIGAL: Great.

So, Anne, could you discuss the evidence?

MEMBER PELLETIER-CAMERON: So, there is quite a bit of evidence in both gynecologic surgery and bladder and ureter injury. However, most of the data does look at hysterectomy data, either laproscopic or open hysterectomy data.

There are several references here. Some of them look at benign gynecologic surgery, but many of them do use hysterectomy data. And so, they are extrapolating from that data which, having participated in these surgeries, is not a big leap, so say that you

are working in the same area. You are operating on the same structures, just in a slightly different way. But the risk isn't there, but I think that is the bit of a gap in the data here, that is not a whole lot of data on -- or I would say there is a moderate amount of data on -- the risk of prolapse surgery injury, but there is a high amount of data if you are looking at the hysterectomy data combined with it.

CO-CHAIR SAIGAL: So, the quality of the data is moderate, because there is no randomized data? It is just observational?

MEMBER PELLETIER-CAMERON: There is some randomized data --

CO-CHAIR SAIGAL: Oh, there is? Okay.

MEMBER PELLETIER-CAMERON: -- specifically looking at hysterectomy. The randomized data is not whether or not to do a cystoscopy, but it is there is randomized data of hysterectomy surgeries where the

complication rate was noted. So, there is no randomized controlled trial that we did a cystoscopy or not --

CO-CHAIR SAIGAL: Right.

MEMBER PELLETIER-CAMERON: -- because that would be unethical.

CO-CHAIR SAIGAL: But for the purposes of the measure, though, there is no randomized data specifically about cystoscopy --

MEMBER PELLETIER-CAMERON: No.

CO-CHAIR SAIGAL: -- use, or no?

MEMBER PELLETIER-CAMERON: No.

CO-CHAIR SAIGAL: So, it is mainly observational. So, moderate quality, but it sounds like there is a lot of evidence or a high amount of evidence, and the direction is all the same.

Anne, you have your concern about value. And they have a paragraph in here about how cost-effectiveness of a cystoscopy shows that it is cost savings above a certain

threshold of ureteral injury, and it is actually universally cost-effective. These are the data in this area. So, there is some that maybe mitigates some of the concern you might have had about this setting at least.

Any other comments about the evidence?

(No response.)

Okay. So, then, let's vote. So, there is moderate-ish level of quality and a lot of data in that direction.

MR. WILLIAMSON: We will now vote on the evidence.

You may begin voting now.

(Vote taken.)

Okay. So, we have 12 yes and 2, no, that the evidence does not meet the guidance.

CO-CHAIR SAIGAL: Okay. And then, Anne, the last part is the gap on this measure.

MEMBER PELLETIER-CAMERON: So,

there are three studies cited referencing the gap. And I did read the description of each of these. A lot of these are survey studies, survey of residents in practice. And one is discussing how many residents get credentialed.

So, the data provided doesn't really provide a lot of hard evidence about whether or not people are doing the cystoscopies or not in practice. There is a survey study.

CO-CHAIR SAIGAL: There is one survey study, basically, about the use of cystoscopy, and it just sort of residents, you said?

MEMBER PELLETIER-CAMERON: Uh-hum.

CO-CHAIR SAIGAL: Okay. So, there is no hard evidence that this is not being done, but I guess my thought -- and I don't know if anyone else wants to comment about this -- that if there is controversy about it, some people aren't doing it, that is very

indirect line of reasoning. I don't know what your gynecologist colleagues have to say about that gap issue.

MEMBER GILL: Yes, I would agree that the data aren't strong, but our opinion or impression or expert consensus would be that there is a large gap and this bears moving forward with it.

CO-CHAIR SAIGAL: Stu?

MEMBER REYNOLDS: Well, the way I interpreted their data here, though, too, is that they were suggesting that there is a big gap between the specialties of urogynecology and female urology almost universally use interoperative cystoscopy, and then maybe the general urologists who they quote as having the much lower number, which suggests, again, that there may really be a big gap in practice.

CO-CHAIR SAIGAL: Okay. Okay. All right.

MEMBER PELLETIER-CAMERON: And I

think the data is not strong, but what data is there is clearly identifying a gap.

CO-CHAIR SAIGAL: Okay. So, then, we will all make our decision about how convincing that is.

MR. WILLIAMSON: We will now vote on the performance gap.

Begin voting now.

(Vote taken.)

We have 1 high, 10 moderate, 3 low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. And then, the last, concept approval. Again, we have seen something like this before. It is looking at an intervention to reduce the morbidity of a surgery. It is very common surgery, and the morbidity is serious if it occurs.

The evidence supporting this is somewhat better than we saw in the previous measure, and the performance gap documentation is moderate at best. But the experts in the

room felt that there is a performance gap, based on their consensus and how they read the evidence.

So, we can vote to approve or not approve.

MR. WILLIAMSON: We will now vote on the overall recommendation of the concept.

Begin voting now.

(Vote taken.)

CO-CHAIR SAIGAL: Okay. So, it is approved.

MR. WILLIAMSON: Yes, we have 14 yes and zero no.

CO-CHAIR SAIGAL: Okay. Any comments for the developers to go home with?

MEMBER PELLETIER-CAMERON: I just had a comment about the use of CPT codes to identify the cystoscopy. I am not a coding expert, but I thought that was bundled with an anterior repair. So, that was my comment.

MEMBER GILL: Right, I agree with the same problem. But when I have checked on

it, it seems to depend on what procedures were done or not. If a sling is included, it seems to be bundled. If it is just a prolapse repair, it may not be. I think we just have to get that straight about the coding, how we are going to identify the numbers.

CO-CHAIR SAIGAL: Some feasibility questions for the developer to think about. Okay.

So, then, now we open this up to NQF member comment about our last wave of concepts for GU.

(No response.)

Okay. Do you want to ask the operator?

MS. WILBON: Is there anyone in the room who has questions or would like to address the Committee on anything, any of the GU concepts? We are kind of wrapping up.

(No response.)

No one? Okay.

So, we will go to the phone.

CO-CHAIR SAIGAL: Okay. So, then, Operator, if you could let anyone listening in make a comment?

THE OPERATOR: Yes. If you have a question or a comment, please press *1.

(No response.)

CO-CHAIR SAIGAL: Okay. Well, then, we get an extra three minutes of break. So, let's come back at 3:15.

MR. AMIN: Well, no, let's come back at three o'clock.

(Laughter.)

MS. WILBON: We are back on schedule.

MR. AMIN: We are 15 minutes ahead of schedule.

CO-CHAIR SAIGAL: Oh, really?

MR. AMIN: So, let's come back at three o'clock.

CO-CHAIR SAIGAL: All right. These guys are pretty tough.

(Laughter.)

So, three o'clock.

(Whereupon, the foregoing matter went off the record at 2:42 p.m. and went back on the record at 3:03 p.m.)

MR. AMIN: Okay. Just as everybody is searching and finding what they need to get started, I will also say that not everything will probably be on the comparison table that you need to get started.

But, essentially, what we want to do -- and again, this is part of the pilot process here. I mean, we normally always have a conversation around harmonization, but, typically, this happens after a fully-specified measure has been endorsed across all four criteria.

You have an opportunity here for a measure that looks at the same measure focus and/or the same target population to address harmonization upfront. So, this gives you an opportunity to look across the measures that you evaluated this morning and this afternoon

to look at how well they work together as a set to understand the various care processes that we are trying to measure.

So, essentially, the nature of the discussion that we want to trigger for right now is to look across the various measures that you approved as concepts and think about which measures could be harmonized across each other, meaning that when we say "harmonize," meaning that the denominator populations are similar, or if there are other considerations for how exclusions are handled, and basically making sure that the measures are giving similar signals based on the target population.

I think, also, as we had a number of concepts that were submitted here, to have a discussion around whether or not you felt that measures should be, more or less, combined in order to get a better signal of the overall care process, which seemed to be, again, the tenor of the conversation that we

were having earlier today around a measure that went down and a recommendation that that be paired with a concept -- I'm sorry -- combined with a concept that was approved to go forward. This is all new language for us or me, us collectively.

So, anyway, what we have put together here is lists of related concepts that are intended to start the conversation around which components and which measures you would like to, first, think that they are logically paired together, paired in the sense that they are trying to measure the same care process for the same population, and then have a discussion of whether they could be combined or whether they need to be harmonized in terms of the way that they are constructed.

So, with that, I will turn it over to Chris and Andy.

CO-CHAIR SAIGAL: Okay. So, then, in regards to this, one of the things you should keep in mind, because basically we are

charged with giving these developers feedback about what they are spending their dollars on. We don't want them to waste their money.

I think we were having a sidebar earlier about the fact that in the past some payers, not Medicare but private payers, have looked askance at measures that were standalone process measures that didn't have any teeth, and they felt they weren't worth using.

I think one way to make a set of process measures more useful to everybody is to combine them, if they are a spectrum of services that could be considered to be, you know, synergistic. So, as we talk about harmonization, we should consider them both across developers, so we don't have five measures that are looking at the same thing, but also within any one measure set if we can have multi-part measures, I think that probably makes sense for everyone at the end of the road, in my view at least.

DR. PACE: Just one thing to consider, because you have seen some multi-part measures where it is essentially standalone measures just combined into one form, but one thing to really consider is, if it is something that a patient should receive both things, assessment and counseling, then the question is measuring all those patients who did receive both things, rather than looking at them separately. So, I think it is looking at it, also, that way, if there are processes that every patient should receive.

CO-CHAIR SAIGAL: Right. A good process of care can really be as long as you want it to be.

All right. So, with that, we will talk about the ones that are on the board here. Am I the one walking us through it? Or how does that work?

MS. WILBON: However you guys are comfortable. We can help, if you want.

CO-CHAIR SAIGAL: Okay. Well,

maybe I will start and then I will open it up. Okay?

So, then, there were four measures that were thematically-related, related to incontinence. And NCQA has a measure that asks about whether the patient says they were asked about incontinence, whether the patient reports a treatment plan, and, also, one that looks at the actual system-level or provider-level; was one provided to the patient in terms of a treatment plan.

And the AUA has a workup measure that says, was the workup appropriate for this patient before they got surgery, and then one that is basically about counseling them about treatment options.

So, some of those are conceptually-related across these four measures.

Just to start us off, I think that the counseling and the treatment plan characterization are sort of conceptually-

related, in my mind. I don't know what people think about it in terms of the differences that are captured, specifically talking about treatment option counseling versus a care plan or basically harmonized.

CO-CHAIR BASKIN: Well, let's be careful with terms here. So, harmonization, if you are talking about C2049 and 2050, that is not harmonization, we are not talking about. We are talking about combining them either into a composite measure or to a combination measure.

Harmonization is when you are going to maintain two separate measures, but you want the populations to match. One is age 50 to 75 and the other one 65 to 75. And the answer is, could they be the same populations? Does that make some sense?

So, I agree with you, though, 2049 and 2050 are both two components of care that should occur on every patient. And is it acceptable, or even meaningful, to measure

each component separately when, in fact, the ultimate outcome is that both components occurred? Now that is combining into a composite --

CO-CHAIR SAIGAL: That is how I meant.

CO-CHAIR BASKIN: Okay. So, harmonization on those two --

CO-CHAIR SAIGAL: I meant harmonization on 0030 and 2050.

CO-CHAIR BASKIN: Okay. Then, we agree, because I was going to talk about harmonization for those two.

CO-CHAIR SAIGAL: Right.

CO-CHAIR BASKIN: So, if you want to do those first --

CO-CHAIR SAIGAL: So, I mean, and this may or may not be a good idea. It is just on the table.

CO-CHAIR BASKIN: Yes.

CO-CHAIR SAIGAL: So, the idea is basically that you have one of these things

that says that you should counsel a woman who is going to have surgery and explain all the treatment options to her. The other says, if you ask this patient, does she say she has a treatment plan? So, those are related ideas. They may not be worth combining. But if you are one individual, you will be measured in several different ways if we have these two things measured.

So, Stu?

MEMBER REYNOLDS: Well, the way that these things are written, they are asking (a) two different populations about two specific conditions. So, the NCQA is men and women, and they are asking about anytime of incontinence, not just stress incontinence. And obviously, the AUA one that we talked about is women, and then those with stress incontinence and those with surgery. So, I don't know how well those -- conceptually, yes, but I don't know how well they really overlap.

CO-CHAIR SAIGAL: Okay. Good.

MEMBER BORDEIANOU: I think, expanding on this point, that NCQA really is raising awareness about the disease amongst PCPs. It seems like that is the point of the measure, to discuss the problem, inform patients of their options, perhaps send them to see experts or specialists, et cetera; whereas, the other one is about appropriate discussion once you see a specialist.

CO-CHAIR SAIGAL: Okay. Good point.

MEMBER TOBIN: Just a question. I can appreciate the value of harmonizing/combining like concepts, but can you even get to the point of saying, yes, combine these if potentially they might have completely different data sources, if you are collecting these measures in completely different ways? So, that is just a question I would pose.

MS. WILBON: Actually, if you

scroll down the table, there are some more rows on the table where we do actually have side-by-sides of the data sources, the level of analysis, because you are right, Judy, that should be part of the consideration on whether or not they are using similar data sources and the level of analysis on whether or not that might have any implications for them further specifying the measure.

CO-CHAIR SAIGAL: Right. One is a survey instrument, and one is CPT II code-related. So, to harmonize them, either the survey would have to be changed or they would have to use CPT II codes in the survey, which wouldn't work.

MEMBER TOBIN: I mean, I can't speak for the other measure developers, but that may be significant for why they submitted a separate measure, if they feel like a different data source needs to be used.

CO-CHAIR SAIGAL: Good point.

Andy, do you have comments about

those two? You are interested in those?

CO-CHAIR BASKIN: Well, you know, I mean, I see why the difference. I mean, this whole thing about the Health Outcomes Survey measure being one group of patients, like females, and yet it is unclear whether there is enough of a male problem that it makes sense to have males, when overwhelmingly the female issue of incontinence is probably much more so than men, and the root causes being entirely different.

MEMBER BORDEIANOU: You just stepped into a big puddle.

(Laughter.)

CO-CHAIR BASKIN: But I see reasons not to harmonize these measures, I guess is the point here. I mean, you know, a member survey is really looking for whole different information about your relationship with your doctor and bringing it up and being able to talk about things. And the other one, the NCQA measure for the PQRS measure is

really measuring a provider and whether they are doing what is appropriate, not just in seeking out the diagnosis, but once the diagnosis is there.

I see good reasons for them to be separate and that they complement each other. They find out different things about the issue. And then, you would respond to that with a different kind of quality improvement activity, depending on which one of these measures showed what. You know what I mean? You wouldn't actually go in the same direction.

CO-CHAIR SAIGAL: Fair enough.

CO-CHAIR BASKIN: So, I think it is reasonable.

CO-CHAIR SAIGAL: Okay.

CO-CHAIR BASKIN: What I do worry about is why they start at age 65. This is just not a problem before the age of 65? I understand that Medicare is probably the impetus behind this, but is there room for

expanding these measures to the larger age groups?

CO-CHAIR SAIGAL: Alayne, you have a comment about the prevalence?

MEMBER MARKLAND: Quickly, yes, two comments. The ratio is usually 2-to-1, women-to-men. So, men definitely have less; it is not inconsequential, and that rate could be as high as 40 percent versus 20 percent, depending on an increase in age. And so, I think it is very relevant to include men in these surveys.

And the second part would be expanding the age a little bit, I think. You know, especially in women the types and treatments may change depending on age as well, the same as men, depending on what type and age, other comorbidities. So, those are valid points.

CO-CHAIR SAIGAL: Okay. Thank you.

So, then, it sounds, from my

understanding of what the group is saying, that probably these two measures should not be harmonized, for reasons that they are different data sources. There is a different intent and different patient populations. So, leave that alone.

Are there any other measures up here, of these four, that people think should be harmonized? I think the merging thing between the two AUA measures, which is not harmonization -- but what is it called again, collapsing? Combining?

(Laughter.)

So, that is one thing on the table. And then, is there anything else that people think, of these four, let's say, that people feel should be brought together, any of the measures that we have reviewed?

I think maybe the mesh measure could be combined with the AUA measures as well that we voted down. It has got, I think, enough going for it. Just maybe it was

constructed in a way that we didn't like. So, maybe the importance of that could be brought into the treatment counseling.

MS. WILBON: I just have a question for Alayne. In terms of expanding the age group, do you have a recommendation on what that age span would be?

MEMBER MARKLAND: Yes, it is a good question. I don't know; the survey itself may be limited to 65 and older. Maybe that is why. I don't know enough about that survey itself.

But I would say I don't have an age cutoff, but menopause is a big factor for urinary symptoms in women. And so, to include a perimenopausal population or a postpartum would also be a very important piece.

CO-CHAIR SAIGAL: Okay.

CO-CHAIR BASKIN: They were pointing out that NCQA is in the room. We all understand the HOS was, obviously, for the Medicare population. So, that is why it is 65

and older. And then, the PQRS was developed for Medicare or was it -- but I guess the point is, is there some consideration, though, of bringing that down into the commercial population in terms of age? Or is that just no one has asked or there doesn't seem to be a calling for that?

MS. WILBON: Can you use the microphone, please?

DR. GIOVANNETTI: The Health Outcomes Survey is Medicare. It does include people under the age of 65, but they are in the disabled Medicare population, and so not exactly the population you were talking about. And we didn't think it was entirely -- there wasn't enough evidence to suggest that they should all be included in this measure.

The PQRS measure was developed originally with AMA, as part of their Geriatrics Work Group. And so, we were focusing on specifically geriatric syndromes. And I believe the argument there for its being

65-plus was that this was the most prevalent in that population and, therefore, the measurement burden was worthwhile in that population, because it is a screening for all people who come into the ambulatory care setting.

So, that is why the ages were set that way. But we are happy to explore, at least on the PQRS side, expanding that age.

CO-CHAIR SAIGAL: Great. Thank you.

Okay. So, then, I am hearing no other suggestions for harmonization or combining that we haven't already discussed. Is that true?

CO-CHAIR BASKIN: The only question is the characterization of how much incontinence somebody has, because the PQRS measure, I forget, it is either you say -- the PQRS is, I think, you either have incontinence or you don't have incontinence. And the Healthy Outcome Survey I think has none, a

little bit, or a lot of it. And the question is, is there some way that we could really be measuring the same gradations in both of these measures.

DR. GIOVANNETTI: So, the reason that the two are different is because in cognitive testing of the question "Do you have urinary incontinence," what we found was that a lot of individuals who did have symptoms of urinary incontinence did not respond "yes" to that question because they didn't feel that it was a problem.

And so, the revision of the question to -- I'm sorry -- no, it was the revision of the question was to try to pull in more people for whom urinary incontinence was maybe not a big problem in their life, but we really wanted to get at that population.

So, the reason that the two are slightly different versus a diagnosis of urinary incontinence versus this kind of small or big is because we were trying to get the

maximum sample we possibly could of people who have symptoms of urinary incontinence. And so, it just had to be slightly different in the way we word it to patients than how we would code it in the charts.

CO-CHAIR BASKIN: But, since then, hasn't NCQA discussed the idea that people with just a little bit of a problem really aren't the same, really are not being treated the same now anymore, and the measure has been changed, I think? Or is it that an accompanying measure, the accompanying NCQA measure has been changed so much, so that the people only with a significant problem, is there further measurement of the provider?

DR. GIOVANNETTI: We discussed that change, but decided not to follow through on that because of the sample size issue. So, as the measure stands now, all individuals are included who report that it is either a small or a big problem.

CO-CHAIR BASKIN: It does sound

like there is some reasonableness to discussing, you know, how you can align the provider measure and the member measure maybe a little bit better since there seems to be this confusion now as to what --

DR. GIOVANNETTI: Well, so the provider level is based off of a diagnosis of urinary incontinence, because our only data source is the ICD-9 codes. And the patient-reported measure, we don't want to exclude it to people diagnosed with urinary incontinence because that implies that they have already had a diagnosis; they have discussed it with their provider. So, we are really also trying to get at that additional population that doesn't have a diagnosis yet.

But I agree. I mean, we can work on that, and there will be changes to this measure.

CO-CHAIR BASKIN: Thanks.

CO-CHAIR SAIGAL: Thank you.

Liliana?

MEMBER BORDEIANOU: I want to say that the only ones that are harmonizing are 2052 and 2063, which are both looking at the use of cystoscopy and procedures for pelvic organ prolapse.

CO-CHAIR SAIGAL: Great idea. We will work it out. Those are two we are going to talk about as well.

So, at this point, I was wondering if we could talk about collapsing the two, 2049 and 2050, and the mesh measure. In my view at least, the denominators and numerators are similar. The intent of the measures is similar. The populations are similar. And I think that it will be a more durable and credible measure over time if they are combined.

I don't know if other people have different opinions.

CO-CHAIR SAIGAL: Aren't the mesh patients already included in the denominator for 2050?

CO-CHAIR SAIGAL: Well, there is no specific numerator saying, "Was a person counseled about mesh." They counsel about the treatment options, including Kegels and everything else. But I think if they added a statement about surgery, including the risks of mesh, that would --

CO-CHAIR BASKIN: So, a subset of those that said, if the surgery is actually going to be mesh, would have an additional requirement that the others don't have, the risk assessment?

MS. WILBON: So, we just brought up the denominators; 2049 and 2050 are the two of the far right columns, and the denominators are essentially the same right now.

CO-CHAIR SAIGAL: Right. And it would keep you away from the issue that we had about only the denominator for the mesh, for people that had mesh surgery. So, I think it would make it a more useful measure in terms of looking at its impact. So, that is what I

would think makes sense.

Does anybody else have -- yes, go ahead.

MEMBER PELLETIER-CAMERON: No, and I agree with you. I mean, just thinking about this clinically, if someone has incontinence and you talk to them about their options, and they say, "Hey, I want to go for Kegels," I am not going to have a big discussion about mesh with them because they have decided to go for Kegels. I don't think the mesh factors into that decision. So, I agree that the denominator of people who actually undergo mesh surgery should have the mesh discussion, but not necessarily everybody.

DR. PACE: But are you saying that they would first elect mesh and then have that discussion? Or wouldn't that be part of the discussion of the treatment options and the pros and cons of treatment options?

MEMBER PELLETIER-CAMERON: I guess maybe if they were discussing surgical

options. Because if someone comes in and you say, "You can lose weight. You can go for pelvic pharmaceutical therapy, or surgery," and they elect to go for physical therapy, it doesn't seem reasonable to go into a lengthy discussion about the risk of surgery with those patients. I don't think that is fair, like that is necessarily measuring quality, because you go over risks of something that they aren't interested in having.

CO-CHAIR SAIGAL: Uh-hum.

Jenifer?

MEMBER LIGHTDALE: I think you could write this generically. So, basically, whatever treatment option you are going to go with, you make sure that all benefits and risks are known, including potentially risk of failure for your Kegels. So, I mean, I think you go ahead and say, whatever your treatment option is, you have disclosed everything.

CO-CHAIR SAIGAL: Probably to make it feasible, it would have to have like

specific points you are looking for. So, I don't know how overspecified we are going to get wit this. But I think mesh is its own special case in this field. And so, I thought that it was an important-enough thing -- it is measure that didn't quite make it, but I think if you changed the language of the numerator so that it is including biofeedback, Kegels, and surgical options, and then put in parentheses that, if surgical options are considered, mesh risks need to be covered. Make that a parenthetical there. That is what I would think.

MEMBER LIGHTDALE: Are medications options here? Yes. So, I mean, also, benefits/risks of the medications need to be discussed. I mean, there is lots of -- and I am sure there are some black-box warnings with some of the --

MEMBER PELLETIER-CAMERON: But there is no medical therapy for stress incontinence.

MEMBER LIGHTDALE: There isn't?

MEMBER PELLETIER-CAMERON: No.

MEMBER LIGHTDALE: Okay.

MEMBER MARKLAND: It is not FDA-approved.

CO-CHAIR SAIGAL: Okay. So, then, how do we proceed after we have this discussion? What happens next?

MS. WILBON: So, if you guys have settled on -- I am not sure if I heard an actual settlement on what, if there is an actual recommendation.

CO-CHAIR SAIGAL: Do we vote?

MS. WILBON: So, this would have to be part of the --

CO-CHAIR SAIGAL: Well, whatever you guys think is the way to go.

MR. AMIN: So, I guess just summarize it. Yes, just a summary, if you want. Let's do that. We are trying things.

CO-CHAIR SAIGAL: All right. So, I will summarize our discussion, and if anyone

thinks I said it wrong, please correct me.

So, we said that we didn't feel like harmonization across the two measures, 0030 and 2050, made sense because they were different data sources and different intents.

We thought that collapsing Measures 2049, 2050, and the mesh one that didn't go forward, 2051, made sense because the populations were similar, and the intent was to ensure a high-quality process of care that resulted in fewer inappropriate surgeries and failed surgeries, and that that would make the measure have more value to a variety of stakeholders. So, we recommend that that would be a good thing to do for those measures.

And then, the last two we want to talk about were these cystoscopic -- I'm a urologist (laughter) -- cystoscopy for pelvic prolapse and cystoscopy for patients having incontinence surgery.

So, there isn't a slide for that,

I don't think.

MEMBER TOBIN: So, can I ask a clarification question? How can you combine a measure that didn't go forward with measures that did? I mean, wouldn't that be excluded, or no?

CO-CHAIR SAIGAL: We already asked about that before we said no, and we were told that we could.

MEMBER TOBIN: Okay.

CO-CHAIR BASKIN: Yes, the measure didn't go forward as is. What we are basically suggesting as a group is that there are aspects of that measure that you can still get the value of that measure in a different way by saying that, when you counsel people about treatment options, if the treatment option you are landing on is potentially mesh, that you additionally have the responsibility to explain about the risks of the mesh. So, you can still get the value of it and not have that separate measure that has other issues.

MR. AMIN: Chris, what would be an easier way to do this is just go ahead and vote on what you have already discussed, and then we can do the other two. Just make sure there is general agreement, and then we will move on.

CO-CHAIR SAIGAL: Okay.

MR. AMIN: Just a hand vote, a quick hand vote.

CO-CHAIR SAIGAL: So, regarding the summary I just made, if everyone thinks that is reasonable, raise your hand.

(Show of hands.)

And if you think it is unreasonable, feel free to raise your hand. My feelings won't be hurt.

(Laughter.)

Okay.

The last two we are going to talk about potentially combining were the ones about cystoscopy. Aliana, you brought that up. So, maybe you could talk about it a

little bit?

MEMBER BORDEIANOU: Well, it seems like the exclusion criteria by the urologists, when we discussed it the first, was pelvic organ prolapse. And then, you know, the second proposal was to include pelvic organ prolapse on the anterior repair. So, why not say all urinary incontinence surgery plus anterior repairs should have cystoscopy, if we are going to go that way?

CO-CHAIR SAIGAL: Are there comments about that? It sounds convincing to me.

MEMBER PELLETIER-CAMERON: I mean, they happen concomitantly so frequently. I don't know the actual numbers, but the rate of a sling surgery with an anterior or apical suspension is very, very high. So, why split hairs over who is in which group?

CO-CHAIR SAIGAL: You may a point when you were reviewing it that, basically, why would you not do it if you were doing

prolapse surgery at the same time, if you are doing a sling. So, it makes sense.

If there are no other comments, then I will summarize that discussion, that we felt that harmonization of 2052 and 2063, which are measures looking at use of cystoscopy after surgery for stress incontinence and cystoscopy after surgery for pelvic prolapse repair, could be combined because they have a similar patient population and the risks and benefits -- the benefits of the measure/intent are the same, that is, to reduce complications of the surgery. And that is basically it.

MS. WILBON: I would just add a point of information, that they both have -- because we did have the side-by-side tables with all the specifications, just a quick overview. The levels of analysis are generally the same. The 2052, which was for cystoscopy during SUI, only specified an individual clinician level of analysis. And

then, 2063 for cystoscopy during prolapse repair, you specified clinician-group-level analysis and the individual clinician. So, very similar level of analysis. And they had the same data sources, administrative claims and paper records.

CO-CHAIR SAIGAL: Yes. Let's vote.

(Show of hands.)

Okay. Are there any other concepts that people feel -- I'm sorry, any dissenting, any no?

Okay. Thanks for bringing that up.

Any other people have ideas about merging, combining?

(No response.)

Okay. And I would like to make one last comment I think that may be relevant. I don't know where this goes on the agenda, but just for the developers to think about the use of CPT II codes and their future in this

whole measurement paradigm and think about ways to specify measures not using those codes conceptually because they may not have a lot of legs in terms of long-term use.

MEMBER BUTT: Could I add to that comment that this sort of, again, goes back to the burden on providers of pulling these CPT II codes? I mean, they are such a difficult thing to do because it is not a simple yes or no. You have to go dig into the chart, often retrospectively, and find out whether the three conditions were met to code it as such.

So, to the extent that things are moving in the e-measures world and the EHR world, I think these developers really need to stay in sync with that and retool these measures so that much of this data should be available from the EHRs.

CO-CHAIR SAIGAL: Okay. So, we have ActiveHealth.

Thank you.

Do we have somebody from

ActiveHealth Management here? Hi. Could you introduce your two measures and give us two or three minutes of time?

DR. WU: I am George Wu from ActiveHealth Management, and this is Dr. Bani Vir.

The first one is GERD patients with alarm symptoms doing an upper GI study or endoscopy. For GERD and alarm symptoms, we are mainly looking at two things, either unintentional weight loss or dysphagia.

As we all know, in the U.S. about 10 to 30 percent of the population has GERD, and it is increasing because of multiple different factors, like stress, obesity -- as we all say, obesity is killing us every single day -- and multiple other factors.

The measure is actually aimed to identify, early identification of complications of GERD; namely, lower esophageal cancers. About 20 years ago, most of the esophageal cancers were squamous cell

in the mid to upper thoracic esophagus. But, nowadays, 1 to 50 percent are in the distal, and GERD plays a major role in that.

Secondly are strictures.

Third is to identify whether PPIs are helpful in treating GERDs.

And fourth is for biopsy opportunities and, as we all know, eosinophilic esophagitis is on the rise right now, and it is part of the differential.

So, that is our first measure.

Our second measure is on chronic liver disease patients and hepatitis A vaccination. We look for patients with chronic hepatitis B and chronic hepatitis C, and see if they ever had the hepatitis A vaccination being done.

Since the introduction of hep A vaccine in 1995, in the United States we see a significant decrease since then. The cases are actually not that much when you look at it. There were only about 2,000 or so

reported cases in 2009. But the underreported cases and the non-reported or the asymptomatic cases could go up to about 20 to 22 thousand in 2009. And especially now, with the CDC's recommendation of screening everybody born between 1945 and 1965, the estimated number of undetected or underreported hepatitis C cases reach about 800,000. So, that is a huge opportunity right there.

So, that is our measure. Any questions?

CO-CHAIR SAIGAL: Okay. John?

MEMBER MORTON: One question about the first measure. How you identify people with alarm symptoms?

DR. WU: So, we use multiple ways. No. 1 is we use claims data, ICD-9s. No. 2 is from our PHR. So, we have a 4-million-user personal health record that actually enters patient symptoms in it, and we actually specify specifically unintentional weight loss and/or dysphagia.

And the third part is through our disease management program, where we also have about 3 or 4 million members. Through that, we obtain this information as well. So, it is a combination of administrative claims data and, also, survey-type data.

DR. VIR: And just to add to that, wherever it is available, we do take in data from the Health Information Exchange.

CO-CHAIR SAIGAL: May I have a question of NQF staff? So, it sounds like there are some proprietary data sources that they are using to measure this measure. If this is a national measure and you don't have access to their various data sources, how would that play out in terms of its being adopted?

DR. PACE: So, I think that is a good question. I guess one of the questions is, do you have those data sources on all the patients or do some patients have one data source and other patients may have two data

sources and other patients three? But we will get back to that in just a second.

So, NQF endorses measures that should be standardized, so that anyone could implement them. This will also come down to feasibility when you actually get to the actual measure, and you may want to comment on it.

But if the specifications are precise enough that anyone could implement if they had EHRs, PHRs, and disease management programs, as long as those data elements are specified so that anyone else could implement them, it could still be NQF-endorsed.

Heidi?

MS. BOSSLEY: I think another good example that we often see is measures that are produced out of a registry, such as FTS or others. Again, the measures are specified precisely, so that anyone else could take that information and implement it. But the data that you see before you comes from that

registry. I think it is very similar to what you are seeing here.

DR. PACE: So, in this case, I think you need to think about the standardization and what happens. Ultimately, when the measure comes in, is it specified so that you know the specifications for all of those different data sources?

CO-CHAIR SAIGAL: Now we will do the phase 2 thing. So, really, let's look at the concept and the quality.

Johannes?

MEMBER KOCH: To that end, dysphagia is, obviously, not a patient report. That is an interpretation of the patient. So, that is a physician taking a good history. And weight loss, my guess is you are asking the patient report, although there would be EHR documentation of actual weight loss. Which is it? Is it the actual documented weight loss or the patient report of weight loss, which they may or may not do at any

particular time?

DR. WU: It is actually a combination of both. There are ICD-9 codes for weight loss per se. Again, this is actually out of more personal practice. Most people probably, if you see someone have intentional weight loss, you would not document weight loss as an ICD-9 code. So, that is how we capture the weight loss from the diagnosis portion, but also from the personal health record portion we have unintentional weight loss per se.

Did that answer your question or no?

MEMBER KOCH: Well, not quite, because in those 4 million people, 200 of them, you know, 100 patients are not getting an upper endoscopy. My guess is from personal experience that there is probably thousands of patients who are getting an upper endoscopy. And the whole question that you are raising is, who is getting the appropriate endoscopy

and what knowledge does a physician have, or should have physician have, at the time that they are deciding do you get one and do you get one? Does the patient have symptoms, right, and how are those documented, right?

DR. WU: That is true.

MEMBER KOCH: So, we are doing lots and lots of endoscopies. There is only a teeny-weeny fraction of patients that you have identified that may not be getting one in some timely fashion, based on data that we know isn't recorded well, which is physicians documenting patient symptoms or patients telling physicians their symptoms, or whether they actually have true weight loss or not.

CO-CHAIR SAIGAL: So, we can have a discussion with ourselves and carry this on

John, are you going to introduce the measure for us in terms of the importance?

MEMBER MORTON: Yes. I think everybody is heard a little bit about the measure. The idea is to take a look

specifically at patients who have reflux with alarm symptoms. Reflux is the most common GI complaint. If you read the Gallup Poll, it is probably two out of three Americans have it. So, it is anywhere from 150 million, maybe even 180 million. There are about 15,000 esophageal cancer cases diagnoses annually. The gentleman is right; there was an increase in the types of esophageal cancers.

The main risks, though, appear to be in the obese and in the male gender populations. I was wondering why those weren't included.

If you look at the citation about alarm symptoms, it is down to essentially two studies. One is a case series, and the other one is from Scandinavia. The idea is to try to identify these people sooner rather than later, before there is disease progression.

The numerator, as you heard already, is people had an upper gastrointestinal study, not specified if it is

EGD or upper GI swallow; the denominator those who are 18 or older with GERD who have these alarm symptoms. I think Johannes has just pointed out, how do we determine who has these symptoms? If they are by documentation by physician, it may not be apparent. I can say that, around weight loss or weight gain, that is generally poorly-documented.

I think that is enough about probably the importance, unless people have questions about it.

CO-CHAIR SAIGAL: I have a question as a urologist.

MEMBER MORTON: Yes?

CO-CHAIR SAIGAL: What you are saying is that it is a very small number of people out of the prevalent population who develop this problem that would need to be identified. So, does that mean that it is like a worthwhile thing in your mind or not worthwhile?

MEMBER MORTON: In my mind, it is

a pretty small yield here, a really, really small yield. If you examine everybody who has got reflux, that is a huge population. If you narrow it down to these alarm symptoms, it becomes a smaller population. But I think that is the problem, is figuring out who these alarm symptoms are.

For the gastroenterologist, the only thing I have found in looking at the data, where there were only two studies about the alarm symptoms, so I don't know how super-specific those are. I know from my practice, male gender is a big one. Being obese is a big one. And so, I would have included those if you are trying to really capture who have got emerging esophageal cancer.

CO-CHAIR BASKIN: So, are you saying that the evidence would support the idea that, if you have GERD and you are either male or obese, that it would be appropriate to do an upper gastrointestinal study? And by the way, they do say that it could be a barium

study or an endoscopy.

MEMBER MORTON: My only point, Andy, is to include that in addition to those alarm symptoms.

CO-CHAIR BASKIN: Yes, but I guess that is my point. The measure here is to measure a population, by the way, not an individual provider, but a population, to see whether a population with GERD and alarm symptoms, are they getting an upper GI study of some sort or another?

So, what is that population? If that population is getting bigger by calling them male and obese, then that population is huge. So, that is what I am trying to figure out.

And even if it is only a small percent with just a couple of symptoms that we were talking about, 2 percent of 150 million people is a big impact problem. So, I am trying to understand what is the real population --

MEMBER MORTON: I don't think that is exactly -- oh, go ahead, Johannes.

MEMBER KOCH: I think the population we are looking at are people who actually have cancer, right? So, it is 15,000, roughly, or less. We are looking at the number of people who did not have an endoscopy with alarm features, having had a history of reflux. This is a teensy-weensy group, right?

MEMBER MORTON: Uh-hum.

MEMBER KOCH: It is patients with cancer.

CO-CHAIR BASKIN: This is just people with alarm symptoms, did they get an upper gastrointestinal study?

MEMBER KOCH: Right, but that is a guideline recommendation. What we are arguing with in GI is that we are doing an endoscopy on everybody, anybody with GERD. So, what we are trying to do is restrict that to people who have had longstanding GERD, 10 years or

more, people who have risk factors of obesity, alcohol, cigarettes, and everybody with alarm features, by the guidelines, should be getting an endoscopy.

So, we are just saying people who have alarm features should be getting an endoscopy. I mean, I don't know that that is that big of a group. We don't have an identified -- out of 4 million, they have identified 100 patients in their group, which is of questionable administrative data, right, because who codes for dysphagia when you are doing an upper endoscopy? Maybe you do or maybe you don't. Who codes for weight loss? That may or may not be documented, right? Everybody is coding for an EGD for GERD. That is part of what we do.

MEMBER MORTON: Just a couple of more points, and these are some of the staff notes that came up in reviewing.

One is that the numerator states it includes patients with at least one gastric

or esophageal cancer diagnosis; the denominator excludes patients with documented gastrointestinal malignancy. Is this construction appropriate?

There was a member comment from America's Health Insurance Plans that it cannot be collected easily, given administrative data; however, it is a good registry measure.

When people reviewed it before this came up, it was pretty split, 3-to-3, in terms of importance.

MEMBER SCHOENFELD: I just want to clarify, we are saying whether or not this is a high impact. I mean, that is a little bit different than performance gap. I mean, high impact, should people who have alarm signs get an upper endoscopy? If that is the way we need to answer this question -- I mean, do we have a high impact from doing an upper endoscopy on people who have GERD plus alarm symptoms? It goes beyond just esophageal

cancer. They may have a stricture, et cetera.

So, is that an impactful thing to do? Okay? Sure, we are going to get to the performance gap. I don't think the performance gap is going to be very big, but we will come to that discussion a little bit later on. But is it something that should be done? Is it going to have a big impact on those people who do have weight loss or dysphagia or anemia, iron-deficiency anemia, who also have GERD?

CO-CHAIR SAIGAL: Phil, I think the issue is, though, not that it is impactful for the individual patient, but then, on a population basis, are you moving the dial in the health of the population of a city? So, if it is 10 people in the city, then maybe it is not high impact. It depends.

MEMBER SCHOENFELD: But I think -- I mean, maybe I missed this in terms of the discussion -- but an important minority, a substantial minority of people with GERD

develop alarm symptoms that aren't going to be treated appropriately unless you do an endoscopy because it is for more than just cancer.

Now I think a little bit further we are going to find that looks like virtually everybody who has documented alarm features actually does get their upper endoscopy, that there might not be much of a performance gap we have to address.

CO-CHAIR SAIGAL: Jenifer?

MEMBER LIGHTDALE: I was just going to ask, though, I think I could see this being more useful for getting at primary care physicians who are seeing patients and treating patients with GERD and are missing the fact that they have alarm symptoms and aren't referring them. So, it is more about referral.

CO-CHAIR SAIGAL: It could be high impact for that? Okay.

MEMBER LIGHTDALE: Yes.

CO-CHAIR SAIGAL: Zahid?

MEMBER BUTT: So, I think that if we define the high impact as large groups of people, then the missing piece in this is the one that you mentioned, which is chronic GERD, because that has the biggest impact on a large number of people, because that is one screening criteria for Barrett's and bad things with chronic reflux.

So, this sort of narrows the denominator substantially because a smaller percentage of them present with alarm symptoms. But if you take that group, it has a very high impact, in my opinion, because you will have a very high percentage of them with something bad there.

But, again, it sort of goes back to several issues with the construct of this measure, which we can discuss later on, because, by definition, many of these people who will get the test will be esophageal cancer, and then you exclude them, because the

exclusion is GI malignancy. It doesn't specify a specific malignancy. So, in a sense, you are sort of excluding your own numerator by diagnosing them.

CO-CHAIR SAIGAL: Is the exclusion a known malignancy? I mean, if you make a diagnosis, you are still excluded?

MEMBER MORTON: I am pretty sure it is a known.

MEMBER BUTT: It doesn't specify when it would be excluded, right?

MEMBER MORTON: I am pretty sure it is known.

CO-CHAIR SAIGAL: Yes.

MEMBER BUTT: Okay. So, it should be previous, prior to this, right?

CO-CHAIR SAIGAL: Right. Yes.

MEMBER BUTT: But that is not specified.

CO-CHAIR SAIGAL: Right.

MEMBER BUTT: But, anyway, the other thing is that this doesn't really have

upper endoscopy as the numerator. It has gastric motility studies. All sorts of things are in here as the upper GI tests. So, if you have any upper GI test, you are in the numerator. I don't know what the relevance of that is.

CO-CHAIR SAIGAL: A lower impact, in your mind?

MEMBER BUTT: I think yes. So, I mean, in terms of the relevance, if someone has a motility gastric emptying study, how is it relevant to a dysphagia patient in GERD?

CO-CHAIR SAIGAL: Comments about that from our GI specialists? No?

MEMBER BUTT: But, you know, that is kind of where I -- so, what I was going to say was that is where we had that earlier discussion, that those are issues with the construct of the measure, but not sort of the importance. I mean, the issue is important, but how it is constructed is some of the problems.

MEMBER SCHOENFELD: Right. I think that will go to the evidence part, yes. I mean, to paraphrase what you said, there is no good data that esophageal motility studies are going to be --

MEMBER BUTT: Well, there is gastric motility.

MEMBER SCHOENFELD: -- gastric motility studies are going to be real helpful in somebody with dysphagia and GERD.

CO-CHAIR SAIGAL: Okay. Liliana?

MEMBER BORDEIANOU: I only wanted to point out that, for some reason, Barrett's is an exclusion criteria, which I found confusing since Barrett's is a precursor.

CO-CHAIR SAIGAL: It is a known pathology before, my understanding is, if you know there is Barrett's, then they are excluded from the study.

MEMBER SCHOENFELD: They have already been evaluated.

CO-CHAIR SAIGAL: Jenifer?

MEMBER LIGHTDALE: One more comment, which is maybe to also to be pro for impact. I guess giving more impact would be to actually notice that you don't actually need to exclude patients 18 years and younger. Actually, the same rule applies; if you have GERD and alarm symptoms, you are going to do an endoscopy.

CO-CHAIR SAIGAL: Great point. We want to bring these aspects of these measures, if we can.

Okay. John, go ahead.

CO-CHAIR SAIGAL: I was just going to say the first part of this is just, is this important or not? That is our first determination. And it comes down to: what is the level of importance? Does it affect a lot of people? If it doesn't affect a lot of people, does it affect a population that is clearly at risk that needs special attention?

I think you could argue that it is the latter, that this is not a huge

population, but we are seeing data to show that that group is increasing, people that get esophageal cancer. So, I think based on the second criteria, a vulnerable population with potential for increase probably meets that criteria.

CO-CHAIR SAIGAL: Okay. Well-said. So, then, maybe we can move to a vote.

I will just summarize real quick the discussion. The comments were that this is probably a small number of people. So, it would be low impact from a population health point of view. However, of the people that this measure would affect, the severity and consequences are high. So, some individuals feel that that makes it high impact.

And there are certainly questions about it is defined and whether it would be made better through different specifications. We have to vote on the measure as it is written now. So, if you think that the way the measure is specified makes it unlikely to

impact the health of that small group of people in whom it is intended to help, that would be a problem. And the age limitations that we mentioned are also of consideration, whether it could be a more impactful or a way to increase its impact.

So, with that introduction, then why don't we vote on the impactfulness of this measure?

MR. WILLIAMSON: We will now vote on the impact. And there are four options: high, moderate, low, or insufficient.

And you may begin voting now.

(Vote taken.)

CO-CHAIR SAIGAL: Let's all vote again. There we go. Okay.

MR. WILLIAMSON: And we have 2 high, 7 moderate, 5 low, and 1 insufficient.

CO-CHAIR SAIGAL: Okay. So, then, we can move on, John, to the evidence that supports the measure.

MEMBER MORTON: So, this gets to

the evidence supporting the measure. What is cited right now for evidence, that the alarm systems -- the alarm systems? (laughter) -- the alarm symptoms will help indicate if there is going to be a problem down the road.

I know there must be more out there, but I just go with what was cited. And we had essentially one case series and we also had one other study out of the Scandinavian journal. So, that was two studies, and I would give them moderate strength in terms of support. I have a feeling there must be more than that, but that is what I was able to see from the evidence.

And then, I guess we are going to discuss the gap in performance after the evidence discussion.

CO-CHAIR SAIGAL: Anyway, Zahid, could you turn off your microphone if you are not using it?

MEMBER BUTT: Oh, I'm sorry.

CO-CHAIR SAIGAL: Thanks.

So, are there other studies/data that people are aware of that have not been brought up specifically to this topic?

MEMBER MORTON: I did forget to mention the AGA did have a technical review. So, that is three.

CO-CHAIR SAIGAL: Okay. So, three studies, and the quality was moderate?

MEMBER MORTON: I would call it moderate at best.

CO-CHAIR SAIGAL: Moderate at best?

MEMBER MORTON: Yes.

CO-CHAIR SAIGAL: Okay. Moderate at best, moderate level of evidence, and the direction was correct.

John, it is a USPSTF recommendation?

MEMBER MORTON: I mean, that is what 1(c)6.3 says. It is not one I read very often. So, I can't help you with that.

CO-CHAIR SAIGAL: 1(c)6.3?

CO-CHAIR BASKIN: It says USPSTF is a grade B.

MEMBER MORTON: The grade here is intermediate strength recommendation.

CO-CHAIR BASKIN: I mean, my point is, if it is USPSTF, then there is a whole body of evidence to support the USPSTF making that recommendation that we may or may not be seeing in this document. That is an evidence-based recommendation that they make at grade B. So, it makes me think that there is evidence that perhaps has not been presented, but it does exist. That body is pretty conservative in their recommendations.

CO-CHAIR SAIGAL: Phil?

MEMBER SCHOENFELD: This isn't really my area of expertise within GI, but having said that, there is a lot more out there on scoping people who have GERD and alarm symptoms like dysphagia than what is listed in here.

CO-CHAIR SAIGAL: Thank you.

Okay. So, there may be data that is out there that we are not seeing. Obviously, if the USPSTF has a B grade, there is a lot of data. The developer put that on here in terms of the grade. And in the document, there is at least a moderate level of evidence to support the quality and quantity supporting the measure.

So, unless there are other comments, we can vote about that.

(No response.)

Let's vote.

MR. WILLIAMSON: We will now vote on the evidence. Again, there are three options.

You may begin voting now.

(Vote taken.)

One more. There you go.

We have 7 yes; 2, no, that the evidence does not meet the guidance, and 6 that insufficient information was submitted to rate.

CO-CHAIR SAIGAL: So, for the people who voted no, can anyone please voice your opinion about what the concern was?

MEMBER MERGUERIAN: The evidence seems to be existing, but they did not provide it to us.

CO-CHAIR SAIGAL: The evidence in the document was not convincing to you? Okay.

Phil?

MEMBER SCHOENFELD: And I think there is an important thing to differentiate here, which is, is there a lot more data out there that they could have included? Yes. Everybody has to make a judgment on their own whether or not that actually rises to the level of saying that what they did cite is inadequate to support the need to do this. And I think that is an important thing just to remind people about.

CO-CHAIR SAIGAL: Sure. Three, you are not supposed to use three just because there is more they could put in there. If it

is convincing enough on its own, you can say yes, and then, later, tell them to put more data in at the end. So, I don't know. I mean, it was a squeaker, but --

MR. WILLIAMSON: It is no.

CO-CHAIR SAIGAL: Oh, is it no? I thought it was yes.

MR. WILLIAMSON: It was 7 yes and 8 no --

CO-CHAIR SAIGAL: All right.

MR. WILLIAMSON: -- split between --

CO-CHAIR BASKIN: Doesn't that mean that there may be other information available?

CO-CHAIR SAIGAL: Well, let me ask the group again, just to be clear, does anyone think they would change their vote, based on the discussion we just had? Or was there some miscommunication about the meaning of vote? Or are we good on our votes?

Can you raise your hand if you

think you are good with your vote?

(Show of hands.)

Yes.

MS. WILBON: So, maybe we should clarify, too, because I am not sure that it was clear. So, option 2 is the evidence that they submitted is insufficient and you don't think that there is anything else out there that they could have found; it just doesn't exist. Option 3 means the information they submitted is insufficient, but there is potentially other data out there that they could have submitted, but what is in the form does not meet the criteria. So, there is some differentiation. Two means it doesn't exist. Three means it exists, but they didn't find it.

DR. PACE: I was just going to say that, either way, it is a no.

MS. WILBON: Yes.

DR. PACE: And then, you can decide on what the next step is.

CO-CHAIR SAIGAL: Yes.

DR. PACE: You know, to talk about the evidence that does exist or --

CO-CHAIR SAIGAL: My only concern was that we are talking about all the other data that is out there, and people might have shorthanded it and said, well, yes, there is other data out there and pressed 3. I may be wrong about that.

Maybe we could just do this: if you think that maybe that we should do it again because of miscommunication, raise your hand.

(Show of hands.)

Two, three, four, five.

So, most people are happy with their votes then. Okay. So, we will leave it.

MEMBER SCHOENFELD: If I understood what John said, he did say for the quantity of evidence it was moderate and for the quality of evidence it was moderate.

CO-CHAIR SAIGAL: Right.

MEMBER MORTON: John will speak for John here.

(Laughter.)

MEMBER SCHOENFELD: Yes. Because if I didn't understand, that is fine.

MEMBER MORTON: Well, to be clear, if I were grading this, you know, like a Cochrane-type deal, this would be poor. You know, we are talking about a case series that is not Level 1 evidence. Okay? That is probably Level 3 at best. And then, we are dealing with a recommendation from a society. Even as August as AGA, it is still a societal recommendation.

And the second one was the best. The third study was out of Scandinavia, three years, a single-site study, though. It wasn't randomized. So, that is why I would put it moderate at best. And we had three studies, mind you. So, that is where I would put it.

CO-CHAIR SAIGAL: Okay. Paul?

MEMBER MERGUERIAN: One of the things regarding the Quality B evidence, I think it says recommends this service. But when you look at the numerator, there are like five or six different studies that are in the numerator, and they did not provide evidence for each one of these studies. So, I am not sure what the service means. Is that the endoscopy? Is that upper GI motility series? Is that an upper GI series? I don't have that data.

CO-CHAIR SAIGAL: You are referring to 1(b), you said?

MEMBER MORTON: I think he is referring to the US Preventive Health --

MEMBER MERGUERIAN: Yes, but, then, when you look at the numerator, there are multiple studies, but I see no evidence, there is no evidence to suggest that each one of these tests -- they are pretty much grouping them all together rather than taking just one test and providing evidence for each

one of these tests specifically.

DR. PACE: So, it is unclear whether there is -- I mean, they give a U.S. Preventive Services Task Force grade, but I don't see the citation for the U.S. Preventive Services Task Force recommendation. Or am I missing it? So, does anyone see a specific U.S. Preventive Services Task Force citation?

MEMBER MORTON: I am sorry, I didn't look that up specifically. I went by that it was in the report. So, I assumed that was correct.

DR. PACE: Right. I understand.

So, maybe we can ask the developer, is there actually a U.S. Preventive Services Task Force recommendation?

DR. WU: So, they used the USPSTF grading. And when you look at the technical review that you pulled up in the PDF over there, page 1397 --

DR. PACE: Okay. So, this is the AGA, and they are using the --

DR. WU: This is the technical review.

DR. PACE: -- terminology that U.S. Preventive Services Task Force uses?

DR. WU: The grading system of the USPSTF grading system.

DR. PACE: Okay. Okay.

CO-CHAIR SAIGAL: That's different. That's different. That is unclear, then. That's different. You are using their grading system, and you are calling the grade fair. Okay.

So, there have been a few post-vote discussion items. I would probably feel better if we voted again, just to be on the safe side. So, is that okay?

All right, let's vote one more time.

Any more questions?

(No response.)

We're good? Okay.

CO-CHAIR BASKIN: But I still have

a comment. So, if that is the grading system used by the -- is this the ASG that used this grading system?

MEMBER SCHOENFELD: AGA.

CO-CHAIR BASKIN: Oh, the AGA? So, I am presuming, then, the AGA has an evidence-based guideline. I mean, you can't have a grading system of grade B if it is not evidence-based. So, once again, there is theoretically an evidence base out there that convinced the AGA to make this recommendation. I mean, you know, they either did or they didn't. I haven't read what studies they cited, but I am going to give credit to the AGA that it is an evidence-based guideline, which makes me think it exists.

CO-CHAIR SAIGAL: Okay. So, that is food for thought for the hopper for the vote.

Let's vote.

MR. WILLIAMSON: We will now revote on the evidence. Again, there are

three options. The first, yes, the body of evidence meets the guidance. The second, no, the evidence does not meet the guidance. And three, that insufficient information was submitted.

So, you may begin now.

(Re-vote taken.)

Did anybody leave the room?

All right. It just counted wrong. Yes.

CO-CHAIR SAIGAL: Okay. So, then --

MR. WILLIAMSON: Yes. So, after re-vote, we have 4 yes; 1, no, that the evidence does not meet the guidance, and 10 insufficient information submitted.

CO-CHAIR SAIGAL: So, the reliability of our voting process needs to be worked on.

(Laughter.)

But we are basically needing to vote now about whether -- I am just kidding --

about whether, despite the fact that there is not a level of evidence to support the measure from an NQF guidance standpoint, we think the measure is important enough to do an override and have it go forward.

MR. AMIN: Actually, Chris, I know it is the last hour here.

CO-CHAIR SAIGAL: Yes.

MR. AMIN: So, because you voted 3, essentially, you are saying that there is an evidence base that exists that was not provided by the developer. So, the question here is actually slightly different. It would not be the exception where there isn't an evidence base. What you are asking here is whether there is a general agreement by the experts in the room that the evidence that does exist that John and others described would meet the quantity, quality, consistency. So that there is the evidence that exists, but it was just not presented, which is slightly different than the exception that the group

has been invoking in the past.

CO-CHAIR SAIGAL: So, I don't know that we have had an explicit discussion of all the evidence exists, but people think there is a lot of it.

So, would anyone who is familiar with it like to give us a comment?

(Laughter.)

MEMBER SCHOENFELD: If you go back to the AGA's technical review on page 1037-38 -- okay, I will speak loudly. No, no, go back up to 1038. Okay.

If you read the wording on that meta-analysis -- maybe you would like to, Chris, if you can see that far? Okay.

The wording there, it says, from where it says, "A recent meta-analysis addressed the specific issue of the utility of alarm signs and symptoms in diagnosing upper gastrointestinal malignancy based on 15 published prospective evaluations encompassing 46,161 patients, 8,669 with one or more alarm

feature, and 150 subsequently found to have gastric or esophageal cancer on endoscopy. Although those investigators concluded that alarm features perform poorly as a diagnostic test, they reported the overall pooled sensitivity and specificity to be 67 percent and 66 percent, respectively."

So, I guess I would just say there that there is a meta-analysis not cited here that says that relying on alarm features to tell you that a patient might need to be screened for cancer is not good. But when you screen everybody with alarm features, you at least do pick up a fair number of cancers.

Everybody can choose how they want to interpret that statement, but it looks like there are a lot of prospective studies that are out there that were not cited.

CO-CHAIR SAIGAL: That is really helpful. Thanks, Phil.

And those numbers are similar to like mammography and PSA testing and those

kinds of things.

MEMBER MORTON: Just one small clarification. There is no mention of this in the United States Preventive Task Force. It is just the grading.

CO-CHAIR SAIGAL: Okay. So, we are all good?

Any other comments about the data that are out there?

(No response.)

Let's vote about whether we think there is high-quality data out there that hasn't been cited yet.

MR. WILLIAMSON: We will now vote on the evidence, the insufficient information provided on the quantity, quality, and consistency. And so, the question is, there is general agreement that the quantity, quality, and consistency of the body of evidence meet the NQF guidance. So, this is a yes/no question.

And you may begin voting now.

(Vote taken.)

And we have 10 yes and 5 no.

CO-CHAIR SAIGAL: Okay. So, that goes back to the developer. The measure stops.

MR. AMIN: No.

CO-CHAIR SAIGAL: We continue? It continues?

MR. AMIN: Yes.

CO-CHAIR SAIGAL: Okay. Good. Okay. Good. Great. Fantastic.

(Laughter.)

John, performance gap?

MEMBER MORTON: The performance gap, I think this is the big question that we all kind of have. We don't doubt that there is not a lot of evidence to support that there are people at risk. The question is, how often are people who are at risk not getting some sort of diagnostic, whether it be an upper GI series or it be an EGD?

They do cite one study where they

believe there is a performance gap of about 33 percent. It is a bit intuitive. They look at a general population and then decide who is at risk for GERD and who has these potential alarm symptoms, and then they see the mismatch between the two.

By that criteria, there is about a 33 percent performance gap, according to them. It is a population of about 4 million people that they did this in, and it was administrative data. But that is the only -- there are no population-based studies.

CO-CHAIR SAIGAL: Any other comments on performance gap?

(No response.)

We have seen this kind of thing before today with -- Johannes?

MEMBER KOCH: Yes, I was going to say that I thought the comment earlier about whether this really applies more to primary care or to gastroenterology is really key. If you have a patient in primary care who has

reflux symptoms and you don't ask them about dysphagia and weight loss, that is an error, right? I mean, that is something that we would want to measure.

This I don't think quite gets at that. And so, I am loathe to just recommend that people do a good history and physical on patients that they see. That seems a little bit mundane for the quality metrics we are trying to achieve.

CO-CHAIR SAIGAL: So, you think a performance gap exists at the primary care level, not the specialist level?

MEMBER KOCH: Right. I mean, I think in specialty it is the opposite because there is lots of data to suggest that we do endoscopies on anybody who has reflux, independent of how long they have it, and there is overutilization of that.

This really is a question, is there underutilization? And there may be under-referral to GI. I don't think there is

an underutilization per se. And it is a slightly different question, I think.

CO-CHAIR SAIGAL: Karen, in terms that this is a population measure, so try to obviate that? Is that what you are saying?

DR. PACE: Right. You know, the question of general versus specialist -- and maybe we want to hear from the developer why they are suggesting this measure is a population-level measure versus at a health plan or clinician level.

DR. WU: To Johannes' comment, from a specialist standpoint, there is zero gap, probably zero gap. But when you have about a thousand primary care physicians out there, I mean a thousand patients per primary care physician out there, you might not remember whether you did the endoscopy or not. You may not remember who had the dysphagia or not. You may not know what the data is coming from, whether someone has the dysphagia, was admitted into the hospital, and you didn't get

that data.

So, this is a way to kind of identify that population that you just forgot about, and mainly geared toward the primary care physician than the specialist.

DR. PACE: Who will use this measure, if it a population, to identify those patients?

DR. WU: Primary care physicians.

DR. PACE: But it is not a primary-care-physician-level measure. You are saying you measure this at a population level, and your example is having 3 million records, or whatever.

CO-CHAIR BASKIN: I think the issue here is using this as a population measure, you are not measuring the performance of an individual physician; you are measuring the performance of a system of care. So, that could be at the level of an ACO-type organization where they are responsible for 50,000 patients, and the answer is that not

all your patients who have alarm symptoms and GERD are getting endoscopies. And then, you would do whatever you do to identify where your issues are. That would be, I think, how this measure is used.

And you may find your issue is not with your gastroenterologist. You may find in investigation that your issue is your primary care doctors. But that is your quality improvement activity.

All this does is identify that you have a problem that you are not capturing all these cases in a population level. That is what the measure is meant to do.

DR. PACE: But I guess I am curious because I made a purposeful choice of not checking like health plan or system. And so, I just want to clarify what your intent is.

DR. VIR: I just want to bring up a point that we had been asked to check off at what level the measure was tested and

specified; whereas, in the past we certainly had this measure marked off. Because this is an endorsed measure by the NQF, we had this marked off at the level of the provider and the health plan and various other levels.

But we were asked to only identify the measure for the level at which it had been tested this time around. And this measure has been tested at the population level. It can certainly be used at the provider level.

DR. PACE: Well, that is the question. I mean, NQF is really trying to be very specific because, especially when you go down in levels of analysis and with the numbers you have seen, to do a provider-level performance measure, the numbers may be too small to actually have reliable -- so, that is why, you know, that is definitely what we are looking for, is that we endorse measures that have been specified and tested, you know, tested at the level for which they are specified.

So, you are right. Thank you.

DR. VIR: Just one other thing. This measure is often used by large organizations, like Accountable Care Organizations. So, they are, as someone brought up the point earlier, they are being looked at across the entire organization, and the organization is able to identify certain providers who have maybe more patients with those issues than others. But it is used for larger organizations.

CO-CHAIR SAIGAL: Thank you.

John?

MEMBER MORTON: Just one comment about the population-based. I think you would have to have a pretty big Accountable Care Organization to make this meaningful and impactful.

If I could just read what was exactly in there, "2.46 million lives are included in the sample population, representing cross-sectional nationwide sample

from our client population. The test of any sort was performed in 260 of 392 eligible patients."

So, you can get an idea of scale. You are talking about 240, 260 patients out of a 2.5 million population. I am not sure exactly how you would implement that on a population basis.

CO-CHAIR SAIGAL: Okay. So, yes, go ahead, Jenifer.

MEMBER LIGHTDALE: I just have two things. If you were going to look at this at the gastroenterologist level -- and I will try to articulate both -- first off, I think gastroenterologists are at risk of interpreting iron-deficiency anemia which can present and, also, with fecal occult blood as only in the colonoscopy, and they forget to do the upper endoscopy. They forget to ask about the GERD. So, that issue is there, and to really remember that they need to be thinking about upper.

And again, I am a pediatric gastroenterologist, but we sort of routinely, if we are going to sedate our kids, we are going to do both, but I know my adult colleagues are not in that same mindset. So, that could play out. There could be some gap that hasn't been identified. I don't know if there is data or I don't know of the data myself.

The other thing, though, I wanted to put on the table is the issue of overutilization because I think that is real. And then, you are going to have a problem not just with negative outliers, or you might not have the negative outliers in gastroenterologists, but you may have positive outliers. I don't know how NQF feels, but, as a gastroenterologist, I start to worry about metrics out there that now are being used against me. So, now I can't get approval to do the procedure that I need to do.

CO-CHAIR SAIGAL: So, then, if I

can sort of summarize, our discussion is about performance gaps specifically. There is, I guess, moderate evidence of a performance gap. There is one big study that John has talked about where there was a small number of people, although there was, whatever, two-thirds of them getting the procedure. So, there is a gap, but it is a relatively-small number of people in that gap.

We had a discussion about many other things related to feasibility, implementation, and measurement level which probably are more related to the stage two of this measure.

But in front of us is a question: is the document as is showing us there is enough of a performance gap for this measure as specified to move it forward?

So, let's vote on that.

MR. WILLIAMSON: We will now vote on the performance gap.

Please begin voting now.

(Vote taken.)

One more. There we go.

CO-CHAIR SAIGAL: Okay.

MR. WILLIAMSON: Zero high, 2 moderate, 13 low, and zero insufficient.

CO-CHAIR SAIGAL: Okay. So, that is not so great.

So, then, the next question is about whether we should approve it. No? We stop here? Okay.

MS. BOSSLEY: So, this is always fun.

(Laughter.)

Because we haven't dealt with this yet in concepts, but this is a maintenance measure. So, there is an opportunity to move a measure into what we call reserve status if it meets all of the other criteria with the exception of the gap.

I don't know that we have thought it through on how this works with the concept, but I think --

CO-CHAIR SAIGAL: Let me comment. The idea is basically that this measure, it is a good job of moving the dial for quality, but there is no room for improvement anymore.

MS. BOSSLEY: Exactly.

CO-CHAIR SAIGAL: But I am not sure that is what we are saying.

MS. BOSSLEY: And that's fine. So, that is why we are raising it as a question.

CO-CHAIR SAIGAL: Because we are reviewing this as a first-time measure. I didn't realize it was a maintenance measure.

MS. BOSSLEY: Right.

CO-CHAIR SAIGAL: And so, apparently, someone else has thought that the evidence was great and that the performance gap was terrific. So, we didn't see that. I am not sure why that is.

No one gave us information about how the implementation of this measure changed the performance gap from when it was first

introduced. So, it is very hard to make that decision.

MS. BOSSLEY: Well, and to be honest, I don't know that you could make a full decision and a recommendation on reserve status until this measure moved into stage two.

CO-CHAIR SAIGAL: Good.

(Laughter.)

MS. BOSSLEY: So, actually, I think what it could be is I actually think you should probably, at a minimum, do a vote on whether you think --

MR. AMIN: Could we vote on importance overall first?

MS. BOSSLEY: Yes. So, I am thinking you should vote to approve or not.

CO-CHAIR SAIGAL: Yes.

MS. BOSSLEY: And then, this is one that we know, if you approve it, it may move forward. And at the time in stage two, you may determine you would like it for

reserve status.

CO-CHAIR SAIGAL: How can you move it forward if you said it wasn't important there is a gap?

MR. AMIN: Well, hold on. Well, let's vote here first. There is criteria for whether it meets the criteria for reserve status, right?

MS. BOSSLEY: Right.

MR. AMIN: And so, they need to vote on that as well. Because I think what we are hearing is --

CO-CHAIR SAIGAL: There is no data about whether high performance is even an actual improvement. That has not been presented to us; I don't think it is in the document.

MEMBER MORTON: I didn't present anything about implementation because there was nothing in there.

MS. BOSSLEY: This is where, typically, if the measure was being reviewed

in the current process, you would have all of that information in front of you. You don't today because we are only looking at the concept.

DR. PACE: They do have the performance on the measure.

MS. BOSSLEY: Right. As it stands now, yes.

CO-CHAIR SAIGAL: We are going to change from prior. So, I don't know. If this is a maintenance measure, what I would suggest is that we are not looking at is -- there was an NQF panel about like measure maintenance, and they specifically looked at what has happened to the population since this measure has come into play. That is maybe a better place for this. Don't you think?

MS. BOSSLEY: Well, you are that group for GI/GU for maintenance.

CO-CHAIR SAIGAL: For maintenance?

MS. BOSSLEY: For maintenance, yes.

CO-CHAIR SAIGAL: Yes.

(Laughter.)

MS. BOSSLEY: So, sorry, but you are.

CO-CHAIR SAIGAL: Is there anything else that we are?

(Laughter.)

MS. BOSSLEY: We will find out throughout the next two days.

CO-CHAIR BASKIN: So, if I may --

MS. BOSSLEY: Go for it.

CO-CHAIR BASKIN: -- you know, at this point it seems that, since this measure had been deemed important in the past, it had been deemed as having a performance gap in the past, we have agreed that it is important up to the performance gap today, and the only way we can assess the performance gap, and whether it has made any difference, is to get to phase 2.

Then, I've got to tell you, it makes sense to send the measure to phase 2

since it is already in place. It is not like we are endorsing a measure that is not being used. It is being used. And all we are saying is we would like to give it the opportunity to go through the feasibility, reliability, validity that it must go through after three years. And it can only be done if we move it on to the next level.

So, my suggestion to those on this Committee is that, based on what I have seen today, and if it is only performance gap that is our issue, that can only be assessed by going to the next level and giving the developers an opportunity to submit the information that we will need.

MS. BOSSLEY: Right. And so, what we would do is bring this forward to you in stage two with a note that it didn't pass the gap now as a maintenance measure.

And I think the other thing we can talk through is what information may be of value for them to bring back in stage two that

might actually look at perhaps the difference in the performance over the last few years. And you can, then, provide your assessment on whether or not it should continue on for reserve status or if you should recommend that endorsement be removed or if it stays as an endorsed measure.

CO-CHAIR SAIGAL: Okay. Jenifer, is that up for you for a reason? Okay.

So, why don't we do a vote, then, on whether we think it should be the reserve status? So, given what we just talked about in terms of its maintenance measurement history, and we don't know anything about how it has changed since it had a debut versus now, can we put it on reserve for reconsideration pending more data about performance data?

So, let's vote on that idea, yes or no.

MR. WILLIAMSON: We will now vote on reserve status. This is a yes-or-no

question.

Please begin voting now.

(Vote taken.)

DR. PACE: This just means that you will continue to evaluate the rest of the criteria. It doesn't mean that you are making any --

CO-CHAIR SAIGAL: Two more. Okay.

MR. WILLIAMSON: Fourteen yes and 1 no.

CO-CHAIR SAIGAL: Is there anything else we have to do?

(No response.)

Okay, next measure.

MR. AMIN: Wait. I think, Chris, before we get there -- sorry -- it sounds like from the discussion here that there is a significant amount of feedback that the group wants to get related to the way this measure is constructed. I think it would be very helpful and good use of the time of the Committee to provide that feedback, so you can

review that when it comes into stage two.

CO-CHAIR SAIGAL: Fair enough. Okay.

So, I think I could summarize the group's comments, and please jump in.

So, the issue No. 1 was, within the evidence basis for the measure, there are existing studies that are supporting the measure that weren't included in the document.

No. 2, in the performance gap issue, we felt there was a small performance gap, given the absolute magnitude of the differences. And we have no information about what has changed in terms of any population health management measurement before the introduction of this measure and today.

So, those are some important things to consider in terms of helping the Committee understand the width of the performance gap for these people. Is that fair?

Zahid?

MEMBER BUTT: Yes, Chris, just to add one more thing is to specify the test and the numerator more precisely.

CO-CHAIR SAIGAL: Thank you for reminding me. So, also, the numerator may be looking at procedures that have no impact on the outcome of interested patients.

Anyone else have any? Paul?

MEMBER MERGUERIAN: Could there be risk stratification, because you talked about the obesity and males? So, is that a measure that should be looked at also?

CO-CHAIR SAIGAL: So, specifying the numerator more specifically to make it more impactful, finding the cases where it should make more of a difference.

Okay. The last measure, which is 0635, chronic liver disease, hep A vaccination.

And, Zahid, can you discuss the importance?

MEMBER BUTT: Okay. Sure. Thank

you.

This is hepatitis A vaccination in patients with chronic liver disease. The measure description is that it is a percentage of adult patients with chronic liver disease who have received hepatitis A vaccine.

The numerator in this case, actually, the denominator is patients age 18 or older who are diagnosed with chronic liver disease, and the numerator is those with chronic liver disease who have received hepatitis A vaccine or who have been tested for immunity in the past. So, this is one of those construct issues that we will probably come back to later on.

But, in terms of the evidence, in terms of the impact first, they present some studies that have been done which show that both chronic liver disease is quite prevalent and common, and the patients with chronic liver disease who develop hepatitis A often are associated with higher rates of fulminant

hepatitis and mortality.

There are several studies that are cited to support this impact and contention. So, I would say that it is moderate to high impact from that standpoint.

CO-CHAIR SAIGAL: Anyone else have comments about hepatitis A vaccination in this population and its impact on the population's health?

(No response.)

Okay. So, then, it sounds like we can have a vote on this concept's impact level, that he feels that it is moderate impact on the population to get them vaccinated to reduce the risk of fulminant hepatic failure.

Yes?

MEMBER BUTT: And mortality.

CO-CHAIR SAIGAL: Mortality.

MEMBER BUTT: Higher mortality.

CO-CHAIR SAIGAL: More mortality.

Okay. Shall we vote?

MS. BOSSLEY: Voting starts now.

(Vote taken.)

MR. WILLIAMSON: We have 7 high and 8 moderate.

CO-CHAIR SAIGAL: Okay, Zahid, quality of evidence.

MEMBER BUTT: Okay. All right. So, now to the evidence that was presented, it is mainly the three guidelines that they have presented, one from CDC and there are two AASLD, the liver folks with their practice guidelines. Those two guidelines, one is specific to hepatitis B and one is specific to hepatitis C, the AASLD guidelines. So, they are a subset of the chronic liver disease population.

The two AASLD guidelines were systematically reviewed. The hepatitis B has a two-three multiple time series, dramatic, uncontrolled experiments and referral.

CO-CHAIR SAIGAL: What is dramatic --

MEMBER BUTT: Yes, it sounds pretty dramatic, doesn't it?

(Laughter.)

And the hepatitis C one has a Class Level 2A, Level C. Weight-of-evidence opinion is in favor of usefulness, efficacy, and the evidence is only consensus, opinion of experts, case studies, and standard of care. So, it sort of comes back to that same guideline issue.

The third one is really by the CDC. Since they are the government, they did not grade it. So, they make the rules.

(Laughter.)

So, all three are pretty consistent in the recommendation that it should be done.

CO-CHAIR SAIGAL: Okay. So, it sounds like it is mainly, then, expert opinion and some observational series supporting a guideline recommendation. Is that right?

MEMBER BUTT: And some dramatic

something.

CO-CHAIR SAIGAL: The dramatic, uncontrolled --

MEMBER BUTT: Yes, dramatic, uncontrolled experiments.

CO-CHAIR SAIGAL: All right. So, then, that would be like, I guess, low to moderate quality.

MEMBER BUTT: Moderate probably.

CO-CHAIR SAIGAL: There are several studies, though, so probably --

MEMBER BUTT: Yes, so there are lots of studies. Again, the AASLD have been graded. So, they at least qualify for a moderate on this. But consistency is high.

CO-CHAIR SAIGAL: The direction --

MEMBER BUTT: All the guidelines are unequivocal about it being done.

CO-CHAIR SAIGAL: Okay. Great.

Any other comments or data people are aware of that is related to this in terms of hep A vaccination?

CO-CHAIR BASKIN: Anyone more familiar than I with the CDC process about evidence base? I mean, honestly, it is not something that I have read up on lately, but, historically, the CDC doesn't make recommendations lightly without some pretty strong evidence. But I am not aware of what their process is.

MEMBER BUTT: Yes. So, again, I am going by what is presented in the submission. Whether the CDC has a lot of information, I didn't go in and check. But, over here, it is mentioned that there is no systematic review through the CDC guideline.

CO-CHAIR SAIGAL: Okay. All right. Then, I guess we could vote on this.

So, let's vote.

MR. WILLIAMSON: We will now vote on the evidence.

Begin voting now.

(Vote taken.)

And we have 13 yes; 1, no, that

the evidence does not meet the guidance, and 1, that insufficient information was submitted.

CO-CHAIR SAIGAL: Great. Okay.

And now, we talk about the gap.

MEMBER BUTT: So, in the gap, they do present a couple of studies that show that the implementation rate remains low. One is a NHANES study which showed that patients with chronic liver disease increased from 13.3 percent to 23 percent over a 10-year period. Similarly, there is a VA study quoted that has a 20.7 percent vaccination rate. So, there does appear to be a gap, and there is some additional registry information that supports that gap concept. They did, again, a sample on their own database and found that there was a 64 percent gap in their population when they used the criteria that they have in this measure.

CO-CHAIR SAIGAL: Okay. Any comments about this performance gap?

(No response.)

All right. Then, I think it is time to vote.

MR. WILLIAMSON: We will now vote on the performance gap.

Please begin.

(Vote taken.)

I think that was a record.

(Laughter.)

We have 11 high, 3 moderate, 1 low, and zero insufficient.

CO-CHAIR SAIGAL: Terrific.

So, then, the last one is --

MR. AMIN: Chris, can I just get some clarification on that last vote? So, my understanding was that there was a general discussion around the fact that there was not -- maybe I am misunderstanding because I just did step in, and I was out of the room for a second -- but that there was not a sufficient gap demonstrated by the material that was presented.

There was information that was described about a 64-percent gap off the patients that were in the sample of 2.4 million, of the 5900 that were identified. So, is that the basis of the fact that it was a high performance --

MEMBER BUTT: No, there were other studies.

MR. AMIN: Okay.

MEMBER BUTT: There are two other studies that show similar low percentages.

MR. AMIN: Okay. Thank you.

CO-CHAIR SAIGAL: So, we are good?

MR. AMIN: Yes.

CO-CHAIR SAIGAL: All right. So, then, this one, basically, is approval of the concept. So, this is, we thought, a moderate-to-high-impact measure on patients who have C, given this vaccination reduces their risk of fulminant hepatic failure and death.

The evidence surrounding it was of moderate grade from different guidelines which

are based on several different types of studies. And we felt the gap was significant, 20 percent performance levels, something like that.

And so now, we can decide if we want to approve the concept in total. So, let's vote.

MEMBER BUTT: So, it is not just hepatitis C. It is all chronic liver disease, inclusively.

CO-CHAIR SAIGAL: Right. Thanks for that correction. All chronic liver disease.

MR. WILLIAMSON: We will now vote on the overall recommendation of the concept.

Please begin.

(Vote taken.)

We have 15 yes and zero no.

MEMBER BUTT: So, Chris, this is now back to sort of the construct issue. In the numerator they define not just people who get hepatitis A vaccination, but those who

have been tested for hepatitis A antibodies. So, not necessarily the result of it, but any test that was done. So, that is sort of inconsistent with the title of this, which says they have received it. So, they make a leap from there that, if you were tested, then someone had the intention to treat you if you were negative for the hepatitis A antibody.

CO-CHAIR SAIGAL: Yes. Right. And so, do you think that is not a valid way to do that?

MEMBER BUTT: I think that I am not sure how you could say that, if you were tested, that that is a priori evidence that you either were immune or received it, if you were not immune.

CO-CHAIR SAIGAL: So, would you like data that looks at the chart, those LOINC codes, and looks at their concordance, and maybe there is a sample of chart extractions between positivity on those antibody tests and inappropriate use of the vaccine, to say that

that assumption is --

MEMBER BUTT: Maybe they, then, should, you know -- well, I just have difficulty in saying that, if a test was done, that is evidence that the patient either was immune or received the vaccination, because those are the two conclusions they draw from a test being done for antibodies.

CO-CHAIR SAIGAL: Right.

MEMBER BUTT: So, I am not sure.

CO-CHAIR BASKIN: So, I mean, this would be an issue of validity.

MEMBER BUTT: Yes, yes.

CO-CHAIR BASKIN: Is it really measuring what it is supposed to measure?

MEMBER BUTT: So, I bring it up for them to address it when they bring it back in stage two.

CO-CHAIR BASKIN: Yes. So, the whole issue would be, for those who were tested and were not shown to be immune, were those patients vaccinated?

MEMBER BUTT: No. Right, right.

CO-CHAIR BASKIN: Right.

MEMBER BUTT: So, this basically just makes the assumption that, if you were tested for hepatitis A antibody, they assume in this measure that you either received it -- you either were positive for the antibody or you actually received it.

CO-CHAIR BASKIN: Right. So, those who didn't have an adequate immunization, right --

MEMBER BUTT: So, I am not sure if you can make that assumption.

CO-CHAIR BASKIN: So, that would be interesting information --

MEMBER BUTT: Something that they would have to --

CO-CHAIR BASKIN: -- if a sample could be obtained to see --

MEMBER BUTT: Right. They would have to somehow prove that that is the case.

CO-CHAIR BASKIN: -- what is the

outcome, right.

MEMBER BUTT: Yes.

CO-CHAIR BASKIN: Right, right.

CO-CHAIR SAIGAL: Any other comments for the developers of this measure?

(No response.)

Okay. Then, any NQF member comments about this last set of activities we have engaged in?

MS. WILBON: In the room, is there anyone who has any comments or questions for the Committee?

(No response.)

CO-CHAIR SAIGAL: Apparently not.

Okay. Then, Operator, are there any public comments?

THE OPERATOR: If you would like to make a comment, please press *1 on the telephone keypad.

CO-CHAIR SAIGAL: I am just amazed that she is listening.

(Laughter.)

(No response.)

THE OPERATOR: And there are no comments at this time.

CO-CHAIR SAIGAL: Thanks.

Good. Okay. So, I think that wraps up for today.

I want to thank everybody for their attention. It is a long a process. I think we have made some progress in getting our rhythm going. So, thank you very much.

And thanks to the NQF staff for keeping us on track.

(Whereupon, at 4:47 p.m., the meeting was adjourned for the day, to reconvene the following day, Tuesday, August 28, 2012.)

**I**

**N**

**U**

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Gastrointestinal Endorsement

Before: NQF

Date: 08-27-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

---

Court Reporter