NATIONAL QUALITY FORUM

+ + + + +

ENDOCRINE MEASURE ENDORSEMENT PROJECT
STANDING COMMITTEE

+ + + + +

WEDNESDAY
FEBRUARY 26, 2014

+ + + + +

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

PRESENT:
WILLIAM GOLDEN, MD, Co-Chair
JAMES ROSENZWEIG, MD, Co-Chair
ROBERT BAILEY, MD, Janssen Scientific Affairs
TRACY BREEN, MD, North Shore-LIJ Health System
WILLIAM CURRY, MD, Penn State College of Medicine, American Academy of Family Physicians
VICKY DUCWORTH, The Boeing Company
JAMES DUDL, MD, Kaiser Permanente
INGRID DUVA, PhD, RN Veterans Health Administration
STARLIN HAYDON-GREATTING, Pharmacy Quality Alliance
ANN KEARNS, MD, PhD, Mayo Clinic
SUE KIRKMAN, University of North Carolina Diabetes Care Center
ANNE LEDDY, MD, American Association of Clinical Endocrinologists
GRACE LEE, MD, Virginia Mason Medical Center

LAURA MAKAROFF, DO, Health Resources Services Administration (HRSA)
ANNA McCOLLISTER-SLIPP, Galileo Analytics
PATRICIA McDERMOTT, RN, Aetna
JANICE MILLER, CRNP, Thomas Jefferson University School of Nursing
CLAUDIA SHWIDE-SLAVIN, American Association of Diabetes Educators
JANET SULLIVAN, MD, Hudson Health Plan
WILLIAM TAYLOR, MD, Beth Israel Deaconess Medical Center, Harvard Medical School

NQF STAFF:
POONAM BAL, Project Analyst
HELEN BURSTIN, MD, Senior Vice President, Performance Measurement
ANN HAMMERSMITH, JD, General Counsel
KAREN JOHNSON, Senior Director, Performance Measurement
KAREN PACE, PhD, Senior Director, Performance Measurement
LINDSEY TIGHE, Senior Project Manager, Performance Measurement

ALSO PRESENT:
MARY BARTON, MD, National Committee for Quality Assurance (NCQA)
KATHY DOMZALSKI, The Joint Commission
DAVID LEE, National Bone Health Alliance
BOB REHM, National Committee for Quality Assurance (NCQA)
ROBERT SAUNDERS, National Committee for Quality Assurance (NCQA)
ETHEL SIRIS, MD, The Joint Commission
ANN WATT, The Joint Commission

TABLE OF CONTENTS

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

8:33 a.m.

CO-CHAIR GOLDEN: Good morning, everyone. Welcome to the Endocrine Steering Committee Meeting.

I'll make a couple of opening comments. I'm Bill Golden. I'm Co-Chair with Jamie Rosenzweig and I am Medical Director at Arkansas Medicaid. I'm also Professor of Medicine and Public Health, University of Arkansas.

Just a couple of my perspectives. This is a big job and I don't know how many of you -- just to help us, how many of you have never been on an NQF Committee before? Okay. Hum. Okay. So, there we are.

This is a big job and it can easily get -- you can easily get lost in some of the rules and nuances, but what we're doing here really determines impact on what people collect. Which is work. Whether or not quality actually improves. Because if you

have a funny measure and people do or do not -- either don't use the measure or don't collect it correctly, they end up not having the impact of making a difference in how care is delivered.

So, there's a lot here. You know, sometimes measures have a vision, but don't have the infrastructure to actually make it happen. So, all of that is really on the table.

And the good news is that because this is a new format, in the old days, if a measure failed because of a technical issue or a specification or something, it was a one-time window and then they were out of luck and apparently with the notion now that we're a standing committee, if we like or the Committee likes the ideas, but the specifications or technical aspect limits the effectiveness of the measure, the developers can come back in six months or a year with a revision. Which is a whole new framework than

used to be.

So, that's an opportunity for us and makes our lives a little easier because we can only approve or disapprove what's written and then what's specified. So, keep that in mind as we move forward.

So, Jamie, do you have some comments?

CO-CHAIR ROSENZWEIG: Sure. I'm Jamie Rosenzweig. I'm an endocrinologist, Director of Diabetes Services at Boston University School of Medicine and also, Associate Professor of Medicine there.

And I've participated on a few NQF committees in the past. The most recent one was on diabetes and cardiovascular disease.

I think we have an awful lot of measures to go through and in two days, I hope we can get through everything in time. So, we're going to be trying to keep things moving as best as we can while giving people enough time to be able to discuss the various issues

related to each of the measures.

But, the whole process is a fairly complex one, but very comprehensive. So, I hope -- I'm looking forward to spending the next couple of days with all of you.

CO-CHAIR GOLDEN: Before we go around the room and have everyone introduce themselves, does NQF staff want to do any ground rules or any information or how do you want to proceed here?

MS. HAMMERSMITH: Hi, everyone. I'm Ann Hammersmith. I'm NQF General Counsel.

What we're going to do is we'll combine the introductions with the disclosures of interest.

It seems that most of you have not served on NQF committees. So, welcome. We're glad to have you here.

I will go through some of the background around disclosures. What we're looking for you to disclose this morning and then we can go around the table.

If you recall, several months ago when you were nominated to the Committee, you should have received an email message to fill out a detailed form regarding your professional activities. We go through those as we are seating the Committee.

Now that you're on the Committee, in the spirit of transparency and openness, we would like you to disclose things that you put on the form or anything that's happened since that's relevant to the work before the Committee. The idea is not to summarize your resume. The idea is to tell your fellow Committee Members and anyone who's listening to the meeting what your interests are that may be relevant to the work before the Committee.

So, we are particularly interested in any consulting activity, research activity, grants that you may have received or speaking engagements, but only if they are relevant to the Committee's work.

Just a few reminders. You sit as an individual. You are here because you are subject matter experts. You don't represent your employer. You don't represent anyone who may have nominated you to the Committee.

The other thing I'd like to remind you of is that our conflict of interest disclosure process is a bit different because we don't ask only about financial interests. Because of the nature of the work that NQF does, we also ask people to disclose if they have done any, for example, work on a committee that has something to do with the subject matter of this Committee even if you weren't paid.

Sometimes that's confusing to people. People will say I have no financial conflicts of interest which is great, but we're also interested in any volunteer activities you have done that may be relevant to the work today.

So, with that, any questions?

I know most of you are new. So, ask if there are any questions before we start. Okay.

We'll go around the table. Tell us who you are, who you're with and if you have anything to disclose and I want to stress just because you disclose something does not mean it is a conflict. The point here is to be open.

So, let's start with the chairs.

CO-CHAIR GOLDEN: All right. So, as I said, I'm a Professor of Medicine and Public Health. I have no financial conflicts. I am on the Executive Committee of the PCPI and I've chaired some of their committees on development of measures. None of them in endocrinology and I do some consulting or potential consulting with General Dynamics in their Performance Measurement Group, but at this point, it's not active in this area either. So, I'd be more or a less a measure consultant.

CO-CHAIR ROSENZWEIG: Yes, I'm on the faculty at Boston University and at Boston Medical Center and I've been chair of several committees at The Endocrine Society that involve performance measures. I was Chair of the Performance Measures Subcommittee for the Endocrine Society as well as I'm now Chair of the Quality Improvement Subcommittee of the Endocrine Society.

I've done some consulting work for some disease management organizations. I'm currently on the Scientific Advisory Board of the Alere Corporation, but I don't have any direct work with them.

MEMBER BREEN: Good morning. I'm Tracy Breen. I'm an Associate Professor of Medicine at the Hofstra North Shore-LIJ School of Medicine. I'm Division Chief of Endocrine there.

I have no financial conflicts of interest to disclose. I serve as a subject matter expert on the Dartmouth High Value

Health Care Collaborative around diabetes and I've also done some collaborations with YMCA organizations in our region around their diabetes prevention program; I think that's the most pertinent.

MEMBER KEARNS: I'm Ann Kearns. I'm from the Mayo Clinic in Rochester. There I serve as the Chair of Quality for Endocrinology.

I don't have any financial conflicts or interests. I've not served on other committees regarding quality measures.

I am in the process of setting up a fracture liaison service at our institution which brings me very close to some of the osteoporosis measures and I'm happy to be here.

MEMBER CURRY: Hi. My name is Bill Curry. I'm a Professor of Family and Community Medicine and also in the Department of Public Health Sciences at Penn State University in Hershey. I'm here at the

invitation of the American Academy of Family Physicians.

In my work at Penn State, I do a lot of quality work, quality measures and a lot of that's around diabetes care. I've done some research with retinopathy and screening for retinopathy and also involved in a project right now looking at the effects of the patient-centered medical home on that outcomes of diabetes care.

MEMBER SHWIDE-SLAVIN: Hi. I'm Claudia Shwide-Slavin. I'm an Advance Practice Registered Dietitian, diabetes educator and I'm representing the American Association of Diabetes Educators. I've done a lot of work with both my organization, the Academy of Nutrition and Dietetics. They've changed their name. Formerly the American Dietetic Association and also with the NCBDE, the licensing board for diabetes educators in development of standards of practice, professional development competencies and I

also do work that I am paid for as a subject expert with the development of education materials with Eli Lilly.

MS. HAMMERSMITH: I'm just going to jump in for a moment and gently remind all of you that you sit as individuals. You're not representing an organization. Thank you.

MEMBER SULLIVAN: Hi. Despite my name tag people call me Jessie. So, my name is Jessie Sullivan and I am the Chief Medical Officer of Hudson Health Plan which is a Medicaid health plan in New York. So, all health plans are measured by HEDIS measures and some of the measures we look at are HEDIS measures. So, in that sense, there is some impact on my life in what happens here, but none of my salary is dependent on that and I have participated on committees for the NQF, for PCPI, for the American Academy of Dermatology and none of the committees that I've participated on were looking at the measures that we're reviewing.

MEMBER KIRKMAN: Hi. I'm Sue Kirkman. I'm an endocrinologist on the faculty at the University of North Carolina.

I have one financial conflict of interest which is that I'm doing a clinical trial for Novo Nordisk where the money goes to my university.

Prior to 15 months ago, I was on staff at the American Diabetes Association and was very involved in their guideline development process. So, may have a little bit of an intellectual, I don't know if it's conflict, but something there.

And while I was at the ADA, I was on several committees with NCQA including their diabetes expert panel and the Clinical Programs Committee that oversaw recognition programs like the Diabetes Recognition Program, the PCMH Programs and so forth. But, it's been more than a year.

MEMBER TAYLOR: Hi. I'm Bill Taylor. I have no relevant conflicts of

interest. I don't think I have an irrelevant conflicts of interest either.

I'm a primary care physician at Beth Israel Deaconess in Boston and I'm on the faculty at Harvard Medical School where I'm an Associate Professor of Population Medicine and an Associate Professor of Medicine and I direct that Primary Care Residency Program at Brigham and Women's Hospital and Harvard Vanguard Medical Associates where I'm also Director of Medical Education.

MEMBER LEE: Hi. I'm Grace Lee. I'm from Virginia Mason Medical Center. I have no financial disclosures.

My research interest previously was grounded in insulin-resistant HIV. When I went to Kaiser Permanente in Northern California, I then became involved with their population-based metrics and published on their hypertension program and currently, I'm at Virginia Mason and have research interest in hospital glycemic control and outpatient

glycemic control.

MEMBER DUCWORTH: Hi. I'm Vicky Ducworth with the Boeing Company and I manage our clinical programs and delivery systems innovation and in a nutshell, that's health systems engineering. I've previously served on CMS' innovations grants as their overview panelist. I've done some consulting primarily in health information technologies.

I am not as accomplished as you all, but if there's a problem, I can find it and I'm pretty good at fixing it. Everything I do is dependent on a measurement. So, happy to be here.

MEMBER MCDERMOTT: Thank you. I'm Patricia McDermott from Aetna. I don't have any conflict of interest that I'm aware of.

I do measures for Aetna for their performance tools. Pay for performance and the like. So, I'm a user of the metrics. So, I'm aware of how -- and I'm very aware of how metrics are constructed and the issues around

the use of metrics with providers. So, that's the expertise I bring to this.

But, as far as conflicts of interests, I don't believe I have any.

MEMBER HAYDON-GREATTING: Hi. I'm Starlin Haydon-Greatting. I'm not on that standing committee roster because I was late to the game. I'm a clinical pharmacist with an emphasis in epidemiology. I means I didn't get a PharmD. I got a Master's in Epidemiology instead. I worked 20 years for Medicaid and did performance measures in the Medicaid populations.

When the State of Illinois drove me crazy, I broke out in shingles and left and went into the private world and I work with self-insured employers in setting up work site diabetes and cardiovascular education programs.

I am part of the American Pharmacist Association, the American Society of Health System Pharmacists and I serve on

the Pharmacy Quality Assurance where we develop measures for adherence and medications.

And I teach at seven -- we have seven pharmacy schools now in the State of Illinois. So, my goal is to educate and create advanced practice pharmacists so that they come out into the world and become part of the team right from the get go and I'm proud to be here. Thank you.

MEMBER MAKAROFF: Hi. I'm Laura Makaroff. I'm a family physician and I work at the Health Resources Services Administration now. I have no relevant financial disclosures that I'm aware of.

My work at HRSA is with the Health Center Program and I work in the office that supports and manages the Quality Measures and Performance Improvement Program for all the health centers that we fund.

So, we are users of NQF measures, but I have nothing to do with measure

development and no financial interests in them. Thank you.

MEMBER MILLER: Good morning. I'm Janice Miller. I'm a nurse practitioner at Thomas Jefferson University in Philadelphia. I'm also a certified diabetes educator. I'm a primary care nurse practitioner for 17 years. In addition to that, I am now an Assistant Professor with the School of Nursing.

I have received and do receive consulting fees from an organization called MyNetDiary as a content expert.

Additionally, I had done some work several years ago on the measure development for some of the cardiovascular measures for a contract organization.

I am just very happy to be part of the Committee and looking forward to working with you all and learning from you all.

MEMBER DUVA: Good morning. I'm Ingrid Duva and I am a quality scholar at the

Atlanta VA with the Veterans Health Administration. I have no conflicts of interest. I have previously served on the ANA Measures Committee for Care Coordination Framework Development and I currently perform some research, I guess you'd call it, with the nurses in our Patient Center Medical Care Homes who are trained to meet the measures that have been developed by implementing different programs to improve diabetes management.

MEMBER LEDDY: I am Anne Leddy. I have done clinical endocrinology in my own office for a very, very, very long time. I am quite interested in all the performance measures because I feel in my heart they're needed and very important.

I have no relevant financial or other conflicts to report.

MEMBER BAILEY: Good morning. My name is Bob Bailey. I work on the Health Economics and Outcomes Research Team at

Janssen Scientific Affairs. I lead diabetes focused projects in the area of health care quality, quality improvement and disparities of care and I'm a nephrologist by training. Was in private practice in nephrology for ten years prior to coming over to Janssen about 11 years ago and I'm an employee of Johnson & Johnson which markets devices and pharmaceuticals in the diabetes base and I'm also a stockholder of Johnson & Johnson.

MEMBER DUDL: Hi. I'm Jim Dudl from Kaiser Permanente. I've worked in -- I am an endocrinologist. I have no financial disclosures. We've worked with performance measures specifically on cardiovascular disease and adherence for many years.

MS. HAMMERSMITH: All right. Thank you very much, everyone.

There are no Committee Members on the phone? No. Okay.

I'm going to give you my final reminder now. With regard to conflict of

interest or bias, if during the Committee meeting you think you may have a conflict of interest or if you think someone else has a conflict of interest, we want you to raise that right away.

You are welcome to do it openly in the meeting. If you don't want to do it that way, you can go to your co-chairs who we'll work with NQF staff or you can go directly to NQF staff. Helen Burstin, our Senior VP for Performance Measurement is sitting right there and you can raise it.

We do not want you sitting there if you're unsure or if you're uncomfortable and not speaking up. It's part of your work as a Committee Member to be mindful of conflicts of interest and bias.

So, if you have any concerns about it, please do speak up.

In that spirit given the disclosures that we've just done, does anyone have any questions of me or anything that you

would like to raise with your fellow Committee Members?

Okay. Thank you.

CO-CHAIR GOLDEN: I guess we will be moving forward. In a little bit, we're going to be doing electronic voting. Correct? Do you want to go over how that works?

MS. BAL: Hello, everybody. So, we will be doing electronic voting and I handed out these little fun notepads to everybody. So, if someone doesn't have one, let me know. Jim may not.

So, basically, each Committee Member will be assigned a keyboard for use during the meeting and you should use the same one everyday. I'll keep track of the numbers and make sure you have the same one.

There is no on and off. It's automatically on once you press this little button right here. I guess it's a little red square and it will turn off automatically once the response is collected.

When you push the button, it will turn green and then no light will show on. If you push the button and then it goes green and then a flashing red, that indicates your battery is low. If it goes just to red, a solid red, that means it's dead and your response did not go through.

You can click the button as many times you want. If you change your mind, go ahead and click it or if you're just not sure if it went through, you can click it again. It won't mess up the system or the count or anything. Every clicker only gets on vote. So, you can click it as much as you want and not have to fear about that.

Basically, the voting cannot start until the timer starts. So, I'm going to do a sample run for everybody. So, you need to -- it's the two screens down at the end. I don't know -- if the voting is not open, it'll always turn red. Yes. So --

CO-CHAIR ROSENZWEIG: Are we

suppose to press send after we hit the button or --

MS. BAL: No. No sending. Just pushing the button.

CO-CHAIR ROSENZWEIG: Okay.

MS. BAL: So, we're going to do a test run. So everybody make sure they understand.

Right now, the screens are in the back. Yes, we'll move them so they're a little more convenient. But, for the sample run, the screens will be in the back and there will be two scales generally. One that's a yes and no which this question is and then the other one will be more of this sort where it's a high, moderate, low and then so on.

So, they'll be rating -- you all received instructions. The rating scale will be on that and then it'll also be on the screens so you can understand it better.

So, let's go ahead and just do one sample run. Once I push the button and you

can see the timer, that's when you can start putting in your answer.

So, right now, you can see the screen's up, the timer's up. So, go ahead and push the button.

Oh, make sure you point at me and not the screen. Sorry.

DR. BURSTIN: The screens are being moved. Sorry.

MS. BAL: Okay. Yes, so, point at me. Yes, if it's goes red, it's bad. Let me know and I can get you a different one.

DR. PACE: Let me explain. It's only if it flashes red. If it -- if you get a red light, it means it's not communicated with the base and just try it again. But, if it's flashing red, then let us know.

MS. BAL: So, you will get 60 seconds for each one and if I get enough responses beforehand, then I'll just stop early. So, we do have 18 responses.

And yes, Jim doesn't have one yet.

I need to give him one. Oh, okay. Then never mind. I'm -- unfortunately, it's been off now, but were you answering? Okay. So, yes, your -- I'll get you a different one.

Great. So, does everybody understand the concept? Is anyone having difficulty with it? Okay.

DR. PACE: You can't tell right now because it's not registering voting. So, the question is whether when you vote it goes green and after that, if you get a flashing red light, then let us know. So.

MS. BAL: Yes, right now, if you push the button, it'll turn red.

I'm sorry. Could you repeat that?

MEMBER HAYDON-GREATTING: So, if you voted and you don't think you voted, you'll get a green flashing --

MS. BAL: No.

DR. PACE: No, that's not what flashing red means.

MEMBER HAYDON-GREATTING: So,

anytime you get a flashing --

DR. PACE: After you get a green. So, let's --

MS. BAL: We'll do one more round.

DR. PACE: Let's do one more.

DR. BURSTIN: This is the hardest part of the meeting.

MS. BAL: So --

CO-CHAIR ROSENZWEIG: Those screens could be moved a little more towards the middle. It would be helpful. I can't really read them.

MS. BAL: Yes.

DR. PACE: What Devon's doing. Yes.

CO-CHAIR ROSENZWEIG: Yes.

DR. PACE: Yes, that's what he's doing. Right. Yes.

MS. BAL: We're shifting them right now. So, basically, if the timer is not on, it will show up red. Because right now, it's not communicating with the system. So,

until the timer's on, anytime you push any button, it's going to show up as red.

So, right now, I'm going to re-push the button and you can see that the timer's on. So, now, you can send in responses. So, we request that everybody send in any response. Doesn't matter what it is as long as it's one through five.

DR. PACE: Right. It has to be one of the numbers that are on this slide. In this case, one through five.

MS. BAL: That means you're not -- you need to point it more towards me. Yes.

MEMBER KIRKMAN: So, if it just blinks green once and then that's it?

MS. BAL: You're good. That means, yes.

MEMBER KIRKMAN: It's in?

MS. BAL: Everything's fine.

MEMBER KIRKMAN: Okay. I guess I have to stare at it while it's --

DR. PACE: You don't have to face

-- do essentially that. Just --

MS. BAL: We'll know.

DR. PACE: -- we'll notice it.

MS. BAL: Yes.

DR. PACE: We'll know if votes aren't registering.

DR. BURSTIN: We'll see the totals. Yes.

MS. BAL: I think we just need to give you a new one. Yours is just busted I think.

DR. BURSTIN: And it doesn't matter how many times you press it, you still just get one vote. So, don't feel concerned if you're hitting it again and again. It's not Chicago.

MS. BAL: So, just one more confirmation. Everybody understands how it works and okay. So, that's pretty much it.

If anybody has questions, you're free to ask during the meeting. Thank you.

DR. BURSTIN: One small. We're

only allowed to have three mikes on at a time. So, just remember to turn off your mike when you're done talking as well or else we'll stop communication.

I just want to add my welcome. I'm Helen Burstin. As Ann mentioned, I head over our Performance Measurement Group here.

So, again, if at any time, any questions, any concerns during the process, please come see me.

And again, really thank you. We recognize this is a lot of work for your volunteer time that very few of us have to give towards these kinds of activities. So, we really do appreciate it.

And just lastly just, you know, you will be hearing from our measure developer colleagues who are lined up on the side here who will be joining us at the table. At the time, we are talking about their measures just, you know, keep in mind there's a lot of work that goes into that process. Before they

get to our door, you know, they've had committees as well who have had these discussions.

It's not really an opportunity to kind of wordsmith or change their measures on the fly. You really kind of give your best thinking about the measure, how useful it could be and again, you know, obviously, be respectful of their intellectual work to date. This really is intended to be a collaborative process with our developers, with all of you, with experts and the multi-stakeholders at the table.

So, thank you.

CO-CHAIR GOLDEN: Developers love their children. Right? So, is that what it is?

The other thing that's useful as a convention is that since it's hard to get people to raise their hands, you get tired, use your card and put it upright if you want to talk. That way we can see that someone is

waiting to be recognized. Otherwise, there would be mild to moderate chaos. So, that would be helpful as well.

And every now and then, you'll get your cards up and we'll say do you want to talk and that kind of thing. It'll help.

Why don't we go over the Karen and Katie to talk about the overview and the project introduction, et cetera.

MS. STREETER: Thank you and good morning. My name's Katie. I'm a project manager here at NQF. Thank you all for coming today. It's nice to finally meet you all after working with your for the past couple of months.

I just wanted to review some of the roles and expectations of the Committee and how we will run the meeting today.

We kind of have a standard script of the expectations that I'm going to read to you. So, as you know, NQF is working to improve committee meetings based on input from

a variety of stakeholders and we've made a few changes to our meeting process.

We recognize that we are fortunate to have the measure developers present and we will be asking them to briefly introduce their measures as they come up for discussion.

Selected work group representatives will then begin to discussion of the measures in relation to the measure evaluation criteria.

We also provided a designated place for developers at the main table during the introduction and discussion of their measures. Here they may more easily respond to questions from the Committee and correct any misunderstandings about their measures during our discussion.

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

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

Committee Members act as a proxy for NQF's membership. As such, this multi-stakeholder group brings varied perspectives, values and priorities to the discussion.

Respect for differences of opinion and collegial interactions among Committee Members and measure developers are expected.

The full Committee meeting agendas are typically quite full. All Committee Members, co-chairs, developers and staff are responsible for insuring that the work of the meeting is completed during the time allotted.

So, ground rules for today's meeting. We ask that all Committee Members

are prepared having reviewed the measures beforehand. We will base -- you will base the evaluation and recommendations on the measure evaluation criteria and guidance. We ask that you all remain engaged in the discussions, attend the meeting at all times except at breaks.

We will be taking a break at 10:15 and I believe it's 2:15 with lunch at 12:30.

We ask that you keep comments concise and focused, avoid dominating a discussion and allow others to contribute and indicate agreement without repeating what has already been said.

And now, Karen Johnson's going to talk about our portfolio.

MS. JOHNSON: Thank you, Katie and good morning, everybody. I'm Karen Johnson. I'm the Senior Director, Office Projects. So, it's nice to see you guys and thank you so much for coming. I haven't got a chance to say hello personally yet, but I will

throughout today and tomorrow.

So, we're doing something a little bit different this time in terms of standing committee. So, Bill has already alluded to this being a pilot and we have transitioned from just calling condition specific committees every three years or so and asking you guys to serve on a standing committee and part of what that will entail is overseeing our portfolio.

So, we have our endocrine portfolio that you guys are now the overseers of, for lack of a better word. It is a new function for us. So, we will all be learning as we go, but we try to put down on this slide some of the responsibilities.

So, what are we thinking when we say you are an overseer of the portfolio?

So, the first is we would like you to provide input as you care to on the relevant measurement frameworks. So, we will be showing you a couple of frameworks. One

for diabetes and one for osteoporosis and these frameworks are designed to help folks think through measure development. So, we'll talk about those in a few minutes, but we will be asking specifically on feedback on the osteoporosis framework because right now, that's a draft.

We would also like for you to know which measures are included in your portfolio and we will be helping you with that and also, ask you to understand the importance to the portfolio and again, as we go through, I think you will understand what we mean by that. But, if you have any questions, you can certainly let us know.

We want you to think about as you consider the portfolio, and again, all of this is stuff that you will have in the back of your mind really, but think about measure standardization and parsimony. So, what we mean by that is it's not helpful a lot of times to have lots of different measures

measuring almost but not quite the same thing. It gets really confusing out there. So, that's what we mean by standardization and also by parsimony.

If there's two measures that are pretty much doing exactly the same thing, why are there two and sometimes there's good reasons to have two, but again, that's something you'll keep in mind not only as you think about the portfolio, but also as you go through the actual evaluation of the measures themselves.

We will use this time and throughout the meeting really to think about gaps in the portfolio. So, as we walk through our portfolios, it'll probably become apparent that there are measures that we don't yet have. So, we will ask you to give us some input on what you think those gaps are and that can go out to the field and have developers think about and take advantage of the good thinking that you guys are doing in

terms of gaps.

We would like you to be aware of other NQF measurement activities for the topic area. So, there's a lot going on at NQF not just in the measured endorsement part of our organization. So, we will give you some information about that so that you also learn what other groups similar to yourselves are thinking about these measures.

We would ask you to be open to external input on the portfolio and you've already had a chance I think to see some of that external input. If you've noticed that in the front matter of the measure submissions, when we had them, we put in some pre-meeting comments that came from outside. So, pretty much the public was invited to make comments on these measures and if we've got any of the comments, we made those available to you.

So, again, that's just so that you are aware of what others out in the world are

thinking about these measures.

We would like you to provide feedback about how the portfolio should evolve. So, that is similar to the gaps discussion, but maybe a little bit different. So, if you have feeling about different ways of measuring or different areas of measurement, that sort of thing, we will give you an opportunity to tell us about that.

And then finally, we would ask you to consider the portfolio when you're evaluating individual measures. So, we will go through the evaluations and we have criteria for you guys to use, but you also will keep in the back of your mind the portfolio and what is really needed to really try to drive quality improvement for interim conditions.

So, let me stop there and see if there's any questions before we go on and look at our portfolio.

Oh, okay and Lindsey just told me

that we have another Committee Member at the table. I'm sorry. I didn't see you come in.

Would you like to introduce yourself?

MEMBER MCCOLLISTER-SLIPP: Okay. Now, here we go. All right. I'm going to break out in song.

My name's Anna McCollister-Slipp. My company is Galileo Analytics, but I'm also here as a Type 1 diabetes patient with complications. So, that's how I got interested in these issues.

MS. JOHNSON: Thank you very much.

Okay. So, to start us off thinking about our endocrine portfolio, right now, the two conditions that we have measures for are diabetes and osteoporosis and you guys are not surprised about that because you've looked at measures for both of those conditions.

Theoretically, we could have measures in this portfolio on thyroid disease,

on metabolic syndrome or on other endocrine conditions. They are in a different color there to show you that right now we do not have measures in those areas.

Okay. Next slide please.

So, this slide and the next really are what Reba calls bringing coals to Newcastle, but just to get us on the same page about diabetes, we know that it is a high mortality condition. It's the seventh leading cause of death in the U.S. right now. Prevalence is more than 25 million and many of those are not diagnosed. Incidents, almost two million new cases per year and it's also a very expensive condition. More than $174 billion per year.

Next slide please.

And this slide is just to remind us all that there are many complications of diabetes including heart disease and heart attack, stroke, high blood pressure, vision impairments, retinopathy and blindness,

chronic kidney disease, potentially ESRD, peripheral neuropathy, peripheral artery disease, poor wound healing and chronic ulceration and then potentially another complication is lower limb amputations.

So, again, those are potential complications and we might be thinking it would be nice to have measures that might look at some of those areas.

So, this slide just gives a quick snapshot. It's not the most up-to-date snapshot, but it's just a quick look at some of the preventive care that is being done in the U.S. and we can see that maybe that preventive care is not as high as we would like those bars to be.

And these kind of reflect some of the measures that we'll be looking at today and tomorrow.

Okay. So, this is our first measurement framework. This is for diabetes and this framework is based on what we at NQF

call our episode of care framework. So, that is a framework that was developed at NQF back in 2008 and really, it is meant to be broadly applicable to different types of conditions and it has a patient-centered focus.

So, you can see how the -- we also call it informally the bubble diagram. But, you start at population at risk and then you go through really the trajectory of disease. So, in this case, phase one is the risk population.

Just a second. She's going to help you out. Yes, we might -- we would get our technical guys to move it. That might work.

Okay. So, the second phase is the evaluation and ongoing management of diabetes and then finally, that third phase that's on the diagram is exacerbation of diabetes and complications treatment.

So, and also what you see on this framework is the idea -- really a couple of

things. Some of the measures, it's kind of hard to say if some of these measures belong in the middle bubble or the third bubble and in a way, that's kind of an academic exercise. It really doesn't matter, but that little set of arrows going around and around in there just indicates that some things just are iterative. You get your care on a regular basis.

What is also shown on this framework is four trajectories indicating different types of diabetes scenarios if you will. So, the first is folks who are in remission or have very tight control. Others who just have the ongoing management. You have a third trajectory that has patients who may go on to have these cardiovascular complications or the forth trajectory, the kidney disease complications.

So, and then also what's pictured here in the framework is things to remind us to think about as we think about measurements.

One is that there is room in the development of measures for patient reported outcomes that reflect diabetes in people with diabetes and there are lots of other issues to think about throughout the episode and I won't read those, but I'm sure you're all very familiar with things like care coordination and access to care and those kinds of issues.

So, let's go to the next slide.

I wanted to walk you through our portfolio. So, I'm walking through again those bubbles. So, the first bubble is population at risk and what this shows you is that we have four measures right now that we have considered as being part of our portfolio under population at risk and what you see from this slide each of the measure numbers has an asterisk by it and that is indicating that we will not -- as an endocrine standing committee, those will not be measures that you will be evaluating. They are evaluated in other projects.

So, one thing that you see there is how we put measures into certain projects or other projects is to some extent arbitrary and we do the best we can.

Obviously, some things could be in two or three different committees. So, the first two, for example, we are looking at in population health. So, they're a more population-based set of measures. So, we'll be looking at those measures in a different project, but they still are under your purview because they are in the endocrine portfolio.

The third and fourth ones there, those are measures relating to diabetes, but they are very narrowly applied to folks in the first one with bipolar disorder and then in the second there, it's schizophrenia or bipolar. So, what that's showing you is that there is some screening and assessment measures that we have, but they are very narrowly focused to this one population of the mentally ill.

Okay. Next slide please.

Most of the measures that we have right now we have placed into phase two, the evaluation and ongoing management and I've split them out into groupings. So, the first one is eye care and you'll recognize the first one, the comprehensive diabetes eye care eye exam measure and that is one that we will be considering later on today.

The next two have to do with diabetic retinopathy and some work around that. Some care processes around that and those again have asterisks. So, those are going to be considered in our HEENT. That's the Head, Eye, Ears, Nose and Throat Project. So, again, a little bit of arbitrariness here, but those are what we have right now on eye care measures for diabetes.

For foot care, we have four measures and all of these are in our work today. We'll be talking about all four of these measures today.

In terms of glucose testing, we will be looking at 0056 today, the HbA1c testing measure and there is another measure that looks at HbA1c as well as LDL cholesterol, but that one is also in a very narrow population. The schizophrenic population. So, that is in our behavioral health project. That's where that one's being looked at.

The next slide.

We have some measures that are directly related to cardiovascular processes. One is LDL screening and appropriate treatment of hypertension. Those you guys will eventually be evaluating. Not in this cycle of the project, but later on and I'm sure you guys are well aware that there have been new guidelines from JNC 8 and AAC/AHA and so, we have purposely pushed those out probably at least until next year so that people can work out any kinks of those guidelines. So, we'll be looking at those a little bit later.

The next two are actually going to be looked at in our cardiovascular project. One measure on kidney disease. You should be familiar with that one because we will be looking at that one today as well and then we have medication measures.

The first one on that list 0541 is a measure that is in a way similar to the three below it, but right now, it is being considered in the safety project. I think because it's a little bit of medication management kind of measure. That one may end up or at least a piece of it may end coming back to you.

So, we're still kind of trying to figure that measure out, but in the meantime, you do have the adherence measures that we'll be talking about tomorrow for statins, ACE/ARBs and oral diabetes agents.

Okay. Next slide.

And then this is what we have for phase three and again, some of those that we

just talked about could have been considered in phase three, but this is what we've said is the phase three. So, we have the poor control and the good control measures and then there's also blood pressure and LDL control measures that just like the other ones that we talked about we'll be pushing those out until at least next year so that we can think about the guidelines that have come out.

We have also a composite measure, optimal diabetes care. That one we have pushed out as well because one of the components of that measure, it's an all or none measure, but one of the components has to do or actually maybe a couple of the components have to do with the LDL and the blood pressure levels. So, again, that one has to be pushed out.

The next two on that list are in orange and that's to signify that you are considering them, but they are new measures that are coming to us this time around. So,

they have not been NQF endorsed yet and that will be what you will decide tomorrow or at least make a recommendation for us. So, they may or may not become part of our portfolio, but they're up for membership in our portfolio.

We have a few outcomes measures. They are complications due to diabetes and those are hospital measures and then some amputation, one amputation measure and an uncontrolled diabetes readmission rate. Those are all a level of analysis as a population. So, again, those are in our -- well, we used to call it population health. I think we're calling it the health and well-being now, but those are being looked at in a different project. But, we do have a few outcomes measures.

And then finally, right now, we do have one resource use measure, relative resource use for people with diabetes and the star there again indicates that that's not

something you guys will have to look at. We have another project that looks specifically at cost and resource use measures. So, they will be evaluating those measures.

Okay. There are several other NQF measurement activities going on that relate to our endocrine measures and the first is the Measure Applications Partnership Diabetes Family of Measures.

So, in case you're not familiar with the Measure Application Partnership or MAP as we call it, it is a public/private partnership that is convened by NQF and it was created for a couple of reasons, but mainly to provide input to the Department of Health and Human Services on the selection of performance measures that will be used in their programs.

So, that one is a statutory requirement that that be done and that group is also, like you, a multi-stakeholder group that considers measures. They do not get into the weeds. So, our group, the endorsement

projects get into the weeds of the measures. The MAP thinks of things a little bit more high level.

So, what they did with their family of measures -- well, let me back up a minute. Not only does the MAP recommend measures for use in Federal programs, but they also try to encourage alignment of measures in the public and private sectors.

So, part of that work, that alignment, they have created different families of measures. So, they actually have a diabetes family of measures and what a family of measures means to the MAP folks is they are sets of related measures and measured gaps that span programs, settings, levels of analysis and populations for specific target areas. In this case, diabetes and they try to indicate the highest priorities per measurements.

That's the gaps and the best available measures in their opinion within

each topic area. So, again, there is a diabetes family of measures that have been decided upon by the MAP people.

And if you're curious, they had some rationale when they were picking measures because there are a few to pick from and generally, they were looking for outcome measures as opposed to if they had the choice, they would prefer outcome over process measures. They noted gaps that -- they didn't really have patient and family engagement measures. But, they did prefer more broadly applicable measures and then that's enough there.

Let's go to the next slide.

Probably the most famous thing that the MAP does is recommend measures for Federal programs and they just went through a set of recommendations. Their report just came out I think a month ago or something like that and basically, what MAP does is for the various Federal programs they either support,

do not support and then they have a conditional support category.

So, I'm not going to read all of these for you. Again, this is just a way of taking input from other folks that have opined about these measures.

There are several of the measures that are in front of you today or that will be in front of you a little bit later that the MAP has not supported for use in their programs. So, the programs specifically listed here are the Physician Compare. That's a public reporting program and then the Value-Based Payment Modifier Program.

And in general, on this slide, I gave you a little bit of their rationale about why they maybe didn't support a particular measure. So, just the first one, they had a preference for outcome measures and also, I just want to make sure that everybody understands that these are the MAP

recommendations and obviously, not everybody agrees that they should or shouldn't have been used in programs. So, you know, there is controversy about the MAP recommendations.

CO-CHAIR GOLDEN: Maybe you can just spend two seconds because that's one of the things that's confusing even to someone like myself bouncing around for awhile. The MAP is not the NQF. The NQF is not the MAP. The MAP -- the NQF has a portfolio of endorsed measures and this indicates a user group and their opinion about using an NQF measure.

So, if the MAP says no, does that continue the endorsement of the NQF measure?

DR. BURSTIN: Yes, it's a good question and it's a little complex and we are increasingly trying to think about how to better integrate those functions because they do feel somewhat detached at the moment.

I think the key issue here is that this was -- and what's not on here is the recommendations for the PQRS Program. Which

is sort of more to the starter set program for a lot of physicians and other clinicians to begin doing quality measurement and which many of these measures are on the list.

I think this was specifically getting to more of the programs that either have a significant financial stake associated with them or the newly emerging Physician Compare. That some of those measures indicating a preference for where they want the portfolio to go.

So, I think Karen's really making this point to give you a sense of since you're talking about the portfolio many of these are where we are right now. This was a sense of a multi-stakeholder group coming forward and saying this is kind of where we want to go to give you a sense of it.

It doesn't mean necessarily that some of these individual measures won't work for a variety of different uses currently and NQF does endorse measures for a variety of

purposes including quality improvement and various accountability programs.

So, I think it was more so in that sense that we're giving you this input as part of the discussion of the portfolio review. What we really want from you as part of this discussion is really quite simple. You've looked at some measures. You have a sense from Karen of what we have in our portfolio.

What's missing? What should we really be trying to incentivize the field to move towards developing as a result of this?

So, this gives you a flavor, for example, of, you know, a clear indication for wanting more outcomes, more composites. The kinds of things we hear a fair amount. Just to kind of put that in context.

Does that help, Bill? Sir.

CO-CHAIR GOLDEN: I think so. It just adds to our complexity, but that gives you a sense.

I guess down the road as we move

along, we may want to keep this piece of paper in front of us because it really has impact on some of the measures and how we move about things I would think. But, I'm just, you know --

DR. BURSTIN: Yes, although one important distinction and some of this -- you know, there's a lot of things in flux at NQF at the moment including this question of whether NQF should ultimately move towards an endorsement decision that's not binary yes/no, but is more nuanced around the particular intended use of the measure. You're going to come against this issue repeatedly today.

At our current point, and you need to act within our current structure and our current rules of the road, we do have binary endorsement. It is yes/no.

This is a much more nuanced interpretation of saying for some of those programs and some people, you know, that the highest impact programs in terms of payment or

public reporting, for example, some the MAP didn't think these measures necessarily rose to that level. They weren't universally considering the boarder -- all broad intended uses of measures.

Ultimately, one question would be -- and we're actually going to do some additional lean work this year to think about how to really better integrate the work of endorsement of MAP. Should the endorsement side put forward clearer recommendations around the science which support this measure for this purpose and the question is how much can the science actually -- what's the underpinning there to say this measure's better for payment, this measure's better for QI and that's where it gets difficult. But, from where we sit right now, we don't yet have that.

So, you need to think about the broadest possible uses of measures which could include quality improvement and some of the

sort of starter programs and again, I think ultimately this process will likely give clearer recommendations. So that when the MAP has to sit down and make these recommendations to the Federal Government about particular programs, hopefully, they'll have additional guidance from these kinds of groups who evaluated really the scientific properties of the measure. We want you to really -- you know, we've grounded the criteria quite clearly into all your materials.

Karen Pace is joining us today as our lead methodologist. Has, you know, worked with our committees and CSAC to try to give you a flow chart to really try to give you a grounding and staying in the science, the criteria, the scientific properties.

The intended uses of the measures will certainly come up as part of the discussion, but again, try to keep in particular this discussion grounded here.

We'll capture some of those other

comments. We'll feed them back to the MAP. We'll feed them back to the developers who are fortunately all here with us today. So, you'll have, you know, real time feedback into those processes.

But, this was really intended to help us think about this is where we are right now. Where do we need to go? How do we try to incentivize the measure development dollars out there to help some of the developers find dollars to actually develop some of these measures that many of you will say you will likely want.

CO-CHAIR GOLDEN: I see we have a couple of questions. Sue.

MEMBER KIRKMAN: So, I don't want to belabor this too much, but can you explain the overlap or are they synonymous of the MAP with PQRS or is the MAP just sort of a group that kind of advises any Federal program whether it's the VA or Medicare?

DR. BURSTIN: Right. At this

point, yes. Sorry to interrupt. At this point --

MEMBER KIRKMAN: And is physician compare? I guess that's my other question.

DR. BURSTIN: Okay. So, at this point, the Measures Application Partnership was specifically asked by the Federal Government to provide input to CMS. So, at this point, it is primarily the CMS programs.

I think there's 30-some odd programs, believe it or not, within CMS, some of us are not surprised by that, where they have to give guidance. Including, for example, the SRD Program, PQRS, across the board. So, it is not unique to PQRS at all.

And what they are asked to do is say here is the set of measures that CMS puts forward on this list affectionately referred to as the MUC list, the Measures Under Consideration, and then the multi-stakeholder groups -- yes, we've loved that nuance there and then the group then tries to think about

does this measure potentially work for this particular program.

So, it's not unique. It's not directly tied to PQRS. That is one of the programs that the Clinician Work Group in particular spent a lot of time talking about just because the volume of measures is so large for that to cover all the various disciplines and specialties.

Is there another question? Jessie.

MEMBER SULLIVAN: It's not a question. It's just a comment.

Since our work group call, I've been thinking so much about something Bill said on the work group call and it just speaks to this contradiction.

In looking at what the MAP says and it looks to me like the MAP is coming from the point of view of what we want for a population or a person with diabetes. That we really want, you know, composite measures. We

want to make sure that everything's done. We want to see the outcomes and I think that makes so much sense from the point of view of an individual.

And I think the contradiction that I have trouble with and I think we're all going to be grappling with is that the measures are not mostly measuring outcomes for an individual or for a population. They're measuring the performance of a physician and so, those are two different things.

So, I think that's where the contradiction is a lot and we're going to be struggling with this and I'm really glad to hear you say that the NQF is looking at maybe not having a binary thing.

But, at this point, we're sort of in the position of wanting to set standards for the care that a person or population will receive based on measures where the physician is the accountable entity. When in order to get to the outcome we need, there's more

involved than the physician.

DR. BURSTIN: Just to build on that, I think that's a good comment, Jessie.

The other really important piece of this in terms of where I think we're all going as well is trying to get to alignment.

So, the last thing you want is the population measure that Bill's using for Medicaid to look different in terms of the science phase compared to the measure that you're using at the physician level.

So, some of this is -- begin saying even if you have a measure in front of you that might be at a physician level or a health plan level, again, because this is the group that's suppose to be the science base for what we do, really look critically at the measurement properties, the evidence. If that works, I think the issue is ultimately thinking about how those measures can move towards aggregation up, for example, to a population level. Even if it's just the

numerator and the denominator kind of gets changed over time, is the science there at least in the way it's being put forward and Bill looks --

CO-CHAIR GOLDEN: So, not to get ourselves into a philosophy class, but that gets back to my original comments about some of our charge. We have measures. We have silos and the ultimate question is does it make a difference and so, that's sort of where we're heading with the MAP and with the NQF.

You know, there's no point in having -- in measuring something if it's just to measure something as opposed to making a difference in care and I think that's ultimately what we're charged with doing is to try to figure out is it just an exercise or does it actually have value in the long run to how people get care and how we exhort people to do things better.

DR. BURSTIN: And just to remind you as you'll go through it again, the four

criteria are, you know, there's a -- they are hierarchical. So, first, you'll deal with importance including evidence. Then scientific acceptability. Then we flipped it. So, then feasibility and ultimately use and usability. So use and usability is one of the four cornerstones here.

But, I think because we're starting from the lens here of the scientific acceptability of those measures, it's hierarchical beginning with evidence and science and testing. So, the use and usability is really important, but it's only really important if you've actually made it through those first few and that's why I think so much of your work today will be around evidence and scientific acceptability of the measures themselves and then assuming that's good, you can move on to feasibility and usability.

But, that hierarchy was intentional to kind of get at that.

MEMBER DUDL: This concept of importance of measuring, the one thing that has escaped me is why we don't have a health plan or whatever level of reporting heart attacks and strokes.

And the reason I mention that is when I give -- have given some lectures on improving diabetes, heart attacks and strokes to a very high level of people in hypertension and they totally miss the need for adding a statin when they're high risk hypertensive and they don't do it and they don't advise it.

I think that it's -- we're missing an element.

Also, there's simplicity. If you say okay, let's just go ahead and let's go right after heart attacks and strokes. We want to drop them 5/10 percent. Which is what we're really after. We're really not after hypertension, blood pressure and lipids. There's real distortion when you go after those subsets and you don't go after -- so, it

just seems like it's funny we don't measure the one thing that we're really trying to go at.

It's that kind of thing I think that if you -- is that what you're talking about that we need to consider?

DR. BURSTIN: Those are really the gaps and there's lots of reasons how difficult that is and people go on and off health plans as our friends from NCQA could certainly tell you. Getting the longitudinality we all know we desperately want from the HRs and other HIEs and other electronic systems would be great.

I mean I think that's why I think some of this discussion is what do we need and then beginning to think about the infrastructure you would then need to get to those measures.

We do have a health care system currently that is using these measures and I think we are responsible as part of the

maintenance process to say do these measures still meet the bar as you're going through your process today.

CO-CHAIR GOLDEN: Did Anna? Okay.

MEMBER MCCOLLISTER-SLIPP: Yes, and as admittedly probably the least scientifically trained person on this panel here as sort of a patient who's nerdy enough to be involved in something like this and occasionally read scientific journals for fun, one question that I have as somebody who's had Type 1 for 28 years. I have all the microvascular complications. There's been lots of effort taking care of myself, doing all the right stuff.

I used to get into very philosophical discussions with my former physician all the time about these issues. He was absolutely excellent and very well known in the field and he always used to say that if, you know, faced with these kinds of measures that he would have fired me as a

client a long time ago because I have very difficult to control diabetes.

So, for me, it's kind of difficult to think about quality measures in a vacuum when you think -- without thinking about what the implications of these quality measures may be both intended and unintended and that's why I already expressed some degree of discomfort with the whole binary thing and I know that's what we're doing here and that's fine. You know, I'm more than happy to do that.

But, it's difficult to divorce those two within this discussion because these are very kind of blunt quality standards that are going out to many, many physicians. Will have real life implications and I think we need to consider that within the context of our discussion when we're deciding even based on lots of data that, you know, 8 is the number.

What happens if you're 8.2 I mean or 8.1 or I mean if you're 7.9, that's fine.

If you're, you know, 8 then that's not I mean. So, I guess that's my primary comment.

DR. BURSTIN: You know, and again, I just want to emphasize how much we really value the patient voice. It's often times -- I've watched enough of these committees over the years to see that it is often the patient who stops a very nerdy conversation in mid-flow about, you know, decimal points on things and just puts it in reality. So, thank you for that.

Again, just to recall, you will get to talk about usability and use and included in there is explicitly a discussion about the positive impact of those measures as well as potential unintended consequences and that was added just in the last couple of years or so explicitly for the fact that people are really increasingly having concerns about unintended consequences of measurement and that needs to be on the table as well.

CO-CHAIR GOLDEN: Well, we're

beginning to develop topics for the 9:00 brandy conversation in the lobby. But, Sue.

MEMBER KIRKMAN: So, I just had a more general question. Although this group kind of looking at existing measures, you know, brought this to mind again and that is if we're reviewing existing measures and we're sort of going through the same process that we would for new measures, how do we deal with issues such as -- I mean there were a couple of measures where I was actually surprised that they were endorsed to begin with because they didn't seem to meet the standards that we're going through now and so, you know, kind of what are the implications of kind of un-endorsing a measure or does that happen or is there sort of a higher bar because it's already out there? I just struggle with that in our work group.

CO-CHAIR GOLDEN: Can I make a comment on that and maybe Jamie can make a comment, too.

Jamie and I have been involved with this. Something like the primordial ooze and there was a time when there was a desperate need for measures and pre-PQRS, there was a demand that, you know, specialists have to have measures and there was a, you know, build as you're flying kind of approach. So, there was some things that were approved because there wasn't anything.

So, just because it exists now doesn't mean it should exist in the future. Because it was, you know, sort of like you go back home. You're not using an Apple 2 Plus any more on your desk. So, you know, it was the best at the time, but may not be the best now.

DR. BURSTIN: In fact -- sorry.

CO-CHAIR ROSENZWEIG: Yes, I would just comment as well that the standard of evidence that's been required for measures has really increased substantially from five/ten years ago.

It used to be that measures basically were derived from guidelines and the guidelines themselves had varying basically evidence standards and that's changed a lot now and because we're actually looking -- in approving these measures here, we have to look at the evidence ourselves to a certain extent.

DR. BURSTIN: Just to build on that comment, again, the criteria have changed significantly over the years. It is a higher bar certainly I think. Certainly around evidence and testing to a certain degree than it was in the past.

As an example, in 2012, we had a hundred measures added to the portfolio and a hundred measures removed from the portfolio.

So, again, I think there is a recognition that, you know, we need new, but also, I think that a countervailing balance to that is there are programs that need measures.

So, I think we also don't want to throw the baby out with the bath water of

something. You know, the common -- I'm sure somebody will say it today. So, I'll be first. Don't let the perfect be the enemy of the good. It's something that will come up a lot as well.

You know, are these helping? To a certain degree, they may not be where we necessarily want to go, but I think not letting perfect be the enemy of the good is an important countervailing balance I think to the raising of the bar.

CO-CHAIR ROSENZWEIG: The other issue, of course, is unintended negative consequences of measures which we have to at least think about. Because I mean sometimes measures will be used for the purpose of paper performance that might be inappropriately used as a base to these measures or physician tiering. Things of that sort which can get very complicated.

MS. JOHNSON: Thank you. What a great discussion right in the middle of these

lists of measures.

Can you go to the next slide please?

Oh, yes, go ahead.

MEMBER BAILEY: Just wanted to make one other comment. In terms of the outcome, any accountability, you have an intermediate outcome. You can hold the current providers whether it's the payer or a physician accountable. It's a longer term outcome. Unfortunately, the retinopathy, cardiovascular disease, those types of complications may have been impacted by care prior to the current entities that are accountable.

MS. JOHNSON: Katie, can you go to our next slide please?

Just so you don't think that the MAP didn't support anything in our portfolio, that's actually not true. They did support several of the measures that you'll be looking at and I'm certainly not going to go through

these lists, but we did want to tell you that there was support and sometimes conditional support for measures and interestingly and I think it was in the staff reviews for the hyper- and hypoglycemia measures that are new.

Again, those have not yet been endorsed. There was conditional support by the MAP for those measures and conditional because it hadn't gone through the in-depth analysis that you're going to look at and also, there was a little bit of concern that those are e-measures. So, something that we'll delve into tomorrow.

And when the MAP does their work, they also identify gaps. So, the gaps that the MAP folks have identified and I think we've already talked about those, they noticed that there's not a lot of measures addressing glycemic control for the complex patients and they didn't see pediatric measures and also not measures looking at the sequelae of diabetes.

Next slide please.

There is additional work that NQF has done. We did a couple of years ago a huge gaps report. So, looking at a lot of different groups of measures and thinking about what might be the gaps in those and again, we're kind of back to the same things that we've already mentioned already. Access to care, patient-centered measures, quality of life, care coordination, communication transitions. So, again, a lot of these things are gaps and these are the ones that were mentioned specifically about diabetes.

And then finally, let me at least tell you that we have a measures pipeline. It was unveiled I think maybe a month ago or a little bit more and the idea of this is to try to start things, some intelligence if you will, about things that developers are working on, things that will be coming down the pipe and we're hoping that they will submit their measures or concepts.

They might not even be fully-developed measures at this point, but we're hoping that they will tell us about them and just so you know, it is new. So, far, we do not have any measures in our pipeline that we know about of diabetes measures. So, nothing to date yet from that source.

Going very quickly into osteoporosis. Again, it is a large problem. High prevalence in the U.S. The main complications are hip fractures and spine fractures, but there are other fragility fractures as well.

And we'll go through the statistics. I'm sure you guys all know that hip fracture and the spine fractures, you know, it is a problem more among women than men, but it is a problem of men and the functional impairment and pain I guess really comes -- and I'm not a clinician. I'm assuming that it comes more from the fractures, but those kind of things do belong

in a portfolio thinking about osteoporosis.

Next slide please.

This graphic is just to show you. What you see there is osteoporosis versus low-bone mass. Just the prevalence by age group. On the right-hand side is women. Left-hand side men. So, it's not something that isn't a problem among men.

Next slide please.

This -- and we don't really have time to go into this, but it is something that we'd like your input on as we go through and you guys are our standing committee. So, we certainly have time to go further in further months. This is our draft episode of care model for osteoporosis.

I neglected to tell you that the model that you saw earlier for diabetes was actually agreed upon by another set of folks who look specifically at diabetes. So, that one was -- we did have a lot of expert input into that model.

This one we pretty much made up ourselves and I think Lindsey did a great job on this one.

So, again, we think that this is a pretty good model, but we will ask you to just at some point -- we may not have time today to go into the weeds of this, but, you know, big picture things. Are these the right things to be thinking about? We have kind of three trajectories there. Are those the right trajectories? Are there other things that should be on our conceptual model as we go through?

Next slide.

This is the one slide that we have for osteoporosis measures. So, the portfolio is very small for osteoporosis. A couple for population at risk. A couple for ongoing management and then a few for post-fracture care.

And what I'm showing you here again as before, the measures with the

asterisk to the side are ones that you will not be evaluating as part of the endocrine group. They belong to other groups.

That first one there is a concept only, but that is -- it actually will live in the GI/GU project and this was -- we tried just looking at concepts. So, this isn't a fully-baked measure yet. It might be at some point. It may come back as a fully-developed measure.

The last three there are new measures that came in in this cycle. So, you guys will be discussing those a little bit later today.

The other osteoporosis measures, we have pushed off until our next cycle. So, we will be looking at those and asking you to evaluate those in the fall.

Okay. Next slide.

In terms of other measurement activities around osteoporosis, same sort of thing. I have what I could find in the MAP

report about the MAP recommendations. I think -- I'm trying to see this. Pretty much the MAP didn't say as much about osteoporosis as diabetes, but there was support for some of the measures, not all of them, and in terms of the gaps report, that one did not look specifically at osteoporosis. So, there was no information on gaps and in our pipeline, we do not have any measures or concepts right now that we know of that are coming along on osteoporosis.

I put this slide in just to make sure that everybody remembers that we do have something called a National Quality Strategy. It is what we think of as our north star really of what things we need to think about in terms of developing and measuring performance. So, we have better care, healthy people, healthy communities and affordable care. That's the triple aim that you hear about and then in the center box are the six priorities.

So, again, as you think about, you know, potential gaps, this is a way to organize different types of measures, you know, affordability measures, patient safety measures, et cetera. Okay.

And so, we've already started this conversation to some point, but we have about 15 minutes I think. Yes, so, here's some questions to consider about the frameworks.

Do they facilitate understanding of improvement opportunities? Because that's what these are suppose to help us do. Think about how we can improve.

And then specifically, any comments about the osteoporosis framework. Again, we might not quite have the time to go into that in depth, but maybe a discussion about why the measures are important. Do they address the -- do the measures that are in our portfolio actually the quality problems or are other types of measures needed?

And then finally as I mentioned

right now, we only have diabetes and osteoporosis. Are there other important conditions and measures for those that we should have in our portfolio and don't?

CO-CHAIR GOLDEN: I think some of that last one could be done at the end of the day or, you know, tomorrow probably for the expansion perhaps.

MS. JOHNSON: Yes, we certainly could.

CO-CHAIR GOLDEN: Yes.

MS. JOHNSON: I think we will have time, but we have about 15 minutes now.

CO-CHAIR GOLDEN: A quick question, then I'll get to the group. I was shocked when I looked at the list a couple of slides back that screening for osteoporosis was not recommended. Can you explain that one? It was very strange looking. I didn't --

MS. JOHNSON: Can you go back, Katie?

CO-CHAIR GOLDEN: Go back two slides I think. Right there. So, MAP, most recent recommendations 0046 do not support for the Medicaid Shared Savings Program. Do not support for Physician Compare. Both of them did not support 0046.

I was just curious. That's a little surprising to me.

MS. JOHNSON: On the first one there for the Medicare Shared Savings Program which that's a program that I'm not really familiar at all with.

DR. BURSTIN: It's the ACO Program.

MS. JOHNSON: Okay. It's the ACO Program. My understanding if I understood right, they only were looking to expand their recommendations for cross-cutting measures. So, I guess those went out and I don't know. That's about the best I can do without really going back and looking, but I can do that for you tonight.

For Physician Compare and the Value-Based Payer Modifier Programs, it's probably -- if I don't have a reason like I had on the other slide, they didn't say specifically other than what they had in quotes does not adequately address current needs of the program and so, I don't know.

Helen, do you recall any more than that?

DR. BURSTIN: Don't have the specifics in front of me and again, there was not significant conversation measure by measure. It was more conceptually just to be clear. So, that I think it wasn't -- they didn't have a specific conversation about that measure and say do not support.

I think it was more so again the idea of wanting more cross-cutting measures particularly for the ACO Program and I think, again, there was a desire, in particular as I recall at the Clinician Work Group, if there were measures of screening that they should

somehow be attached to a follow-up action.

So, I think things that were pure screening without a follow-up action were ones that were not in general preferred. So, I think that was sort of caught in this net as opposed to being anything about particular scientific issues around the measure itself.

CO-CHAIR GOLDEN: Sue.

MEMBER KIRKMAN: Yes, I just had a comment about the osteoporosis episode of care and that is that the focus on prevention of fractures is only in the box for the relatively healthy adult and then once someone's had a fracture, it falls out and since, you know, having had a fracture is the biggest risk factor for a subsequent fracture, I just think that that needs to not fall out. You know, once you've had a fracture, prevention of fractures should be really important and is probably where the best evidence is and the best bang for the buck. So, that was just my comment.

CO-CHAIR GOLDEN: Other comments or questions. Let's go to the last slide again. I guess you can remind folks. Well, good.

MEMBER KIRKMAN: I mean I guess this is going to wait until tomorrow, the discussion about other measures, but I would hope that we could also maybe discuss some measures of overuse because I do think in endocrinology there are, you know, sort of overuse. I'm thinking thyroid nodules and ultrasounds and, you know, there's even emerging evidence that perhaps thyroid cancers are being over diagnosed and treated. So, just something to think about.

CO-CHAIR GOLDEN: As an aside, I'm working with a committee now on radiology measures looking exactly at that. So, there are a couple in the pipeline.

CO-CHAIR ROSENZWEIG: I think that's a very good point. A lot of the data related to the thyroid nodules especially.

MEMBER SHWIDE-SLAVIN: One of the other endocrine conditions, I was wondering why it wasn't included, is pre-diabetes. That's huge and it's not anywhere.

DR. BURSTIN: It' probably in more of our population health focused where more of sort of pre-condition measures are focused, but I'm not even sure there actually is a measure yet on -- I think there's one newly proposed on diabetes screening at a population level. Good point though.

MEMBER MCDERMOTT: I mean there's huge efforts for metabolic syndrome even from health plans. If you are in any of the big ones, you get credit on your premium now for doing that kind of screening no matter what age you are and so forth and so on.

So, I think that that's a very good topic and even the definition of qualifying for a metabolic syndrome and how that's working I think is a very interesting topic to explore at some point.

CO-CHAIR ROSENZWEIG: Is metabolic syndrome within the purview of this committee or is it more in the cardiological sphere?

DR. BURSTIN: It crosses it. It doesn't matter. That's the whole point of having standing committees. You guys can work collaboratively over time and figure out just what needs to get done and where it could live is less of an issue.

MEMBER MCCOLLISTER-SLIPP: And maybe I'm missing something, but why is thyroid not addressed? Was it? Did I miss something?

DR. BURSTIN: Thyroid's in this portfolio. We just have very few measures and they're not yet up for maintenance.

MS. JOHNSON: Right. We don't have any.

DR. BURSTIN: We had a few. But, okay. I thought in the past.

MS. JOHNSON: Yes.

DR. BURSTIN: Yes, very few

measures.

MEMBER DUCWORTH: It would be a better -- so, I think a better understanding of how certain thresholds are established on clinical outcomes measures.

My understanding, and it could be naive is, I guess, the body of evidence. They just perform a series of retrospective analyses and they determine a population mean and what's the most desirable, I guess, target. For instance, like an 8.0 under for A1c.

But, are they looking at the overall population mean? Are they then taking a subpopulation and then targeting that? The ideal for that ideal population. How are they determining that threshold? Because that tells us a lot about, you know, what is desirable in establishing these metrics.

It goes back to what Anna was suggesting. Do we target an 8.0 and under or an 8.1 or 8.2?

The scientific community, how are they identifying that threshold?

DR. BURSTIN: We'll shortly get to that conversation where those measures come up.

MEMBER DUCWORTH: Okay. Yes. Thanks.

DR. BURSTIN: We've got lots of folks at the table to help. But, that's -- those are, you know, important evidence questions.

CO-CHAIR GOLDEN: What you're saying is the diabetic is not a diabetic is not a diabetic?

MEMBER DUCWORTH: Well, I just don't -- okay. So, for instance, I'm from the private sector. Right. Am I going to save my organization $200 million by getting everyone at 7.8 and below or 7.5 and below? You know, the cost savings would be workplace improvements, improved quality of life for our members.

But, I want to know the science. Like how are they approaching that? Because if I don't like the science, I'll just have a more aggressive target.

CO-CHAIR GOLDEN: Thank you and as we move along, if you can move your mike a little closer. You're a little more difficult to hear than some. That's great. Super.

MEMBER DUCWORTH: Oh, sorry. Thanks. That's it. I'll wait for the future conversations.

MEMBER BAILEY: I also just wanted to raise the topic of BMI because currently it appears to be identifying the population at risk, but then when you look at the population that has diagnosed diabetes specifically Type 2 and the evidence-based guidelines recommend weight loss or increased exercise in that population. I'd advocate for also having a measure in that population that's also diagnosed as well or at least, development of a plan or evidence of a plan that has been

discussed.

MEMBER DUVA: So, I feel like we're jumping back and forth on topics related to this slide, but if we can go back to your framework if you're looking for feedback on the osteoporosis and we've not talking about it again later, I think it was Sue that had a great comment about the middle box losing the prevention of fractions.

And I also wanted to throw in there you're looking at this sort of model, but the screening is so important in so many of our populations depending on where you're at. The majority of people could be screened at risk.

So, I think going on with Sue's comment in the next box what you're really interested in maybe is injury prevention. Because once you get to a certain population perhaps the elderly or whomever it may be, the screening, you know, it's kind of like an 80/20. You might have 80 percent of your

people screening positive. So, then what you're really interested in is the injury prevention and it affects a lot of the processes.

CO-CHAIR GOLDEN: You know, this kind of a conversation's one of those things where you'll be walking your dog or carrying out the garbage, you know, you'll have a great thought come to you. You might want to have this sort of as an ongoing share point where people can submit these ideas over time as part of our activities since there's, you know, no point in having it in a five-minute window. We might as well keep our creative ideas available for future use.

Yes, sir.

MEMBER TAYLOR: I'm not sure if this is the moment. This will come up repeatedly as we speak, but we're stuck in a sort of difficult paradigm that the diabetes and osteoporosis measures both exemplify really well.

Which is we have people at risk based on some continuous variable that's a poor predictor of what's going to happen, but the best we have. Like bone mineral density or hemoglobin A1c and when we take the population and that distribution and arbitrarily cast a line somewhere. We're going to have that problem that Anna said so eloquently about people who are close to that line.

What we want clinicians to do, since these end up being measures to help encourage physicians to do what's right, is to do things when there's more benefit than harm from the patient's perspective and if you look at the distribution, wherever you draw the line, the most people in the distribution that we care about will be clustered right around the place you draw the line wherever you draw it. Right.

And then you'll have the problem with, you know, it's the person with the A1c

of 8.1 somehow needs an intervention. Well, you know, if the intervention has some risk going from 8.1 to 8, might do more harm that good. Right.

Then you get into the issue of patient centeredness and values and how you make that distinction and those sort of problems pervade this whole approach where we define a disease by taking a distribution, drawing a line and saying on this side of it, you have the disease and on that side of it, you're okay.

CO-CHAIR GOLDEN: So, welcome to the second half hour of our brandy conversation.

You know, you get into the issue of if everybody has a similar population, you know, it's a normative process and you can compare rates as opposed to the individual position everyone's going to have a variant.

But, your point is well taken and it gets to her point about is everybody the

same. So, but that's a good point for coffee maybe.

Maybe we can say that for the last time during the next two days we're on time as far as schedule if -- I'm sorry, Tracy. Didn't see you.

MEMBER BREEN: Okay. Just one comment and again this is kind of for very heavy brandy later, but looking at these measures as we slice it across, what we're really, I think, looking for as clinicians is the delta. Right? The delta of taking someone with an A1c of 10 and moving them to 8.5 and the risk reduction that happens there or the delta of taking someone with a LDL of 130 and getting them to 105 and the risk reduction.

And I know that's beyond the scope of what we're talking about, but that's really what we're looking at and that's what the doctors are looking at in their practices when they're managing complicated high-risk

patients. They want credit for the delta.

So, I just want to throw that out there as a goal much later.

MEMBER MILLER: Along that same line and in the vein, pardon the pun, of having a delta, we have measures that we're going to be discussing about diabetes education regarding foot exams. I'm very interested in what measures we have. I haven't seen any measures about diabetes education itself.

And we know that the majority of care happens outside of the office. We know that patients are not always aggressive in seeking out care and they -- or I'm sorry. Of seeking out education and the information they get is from mostly, you know, precariously reliable sources.

So, I'm very interested in what measures we have or are being considered for diabetes education locations.

CO-CHAIR ROSENZWEIG: That's an

interesting point because it's been very difficult for measures developers to come up with the specific criteria for judging whether or not a person's received diabetes education or not.

Interestingly enough, the recent NCQA AMA PCPI group that's been developing measures has come out now with an education measure, but for years, we were working on developing diabetes related education measures and when we came up to the evidence-based issues and how to actually define education, they got shot down. So, it's been a very difficult issue.

But, we're sort of in the situation really of considering measures that have been developed by others. I mean I don't think our mandate is to develop measures. We can suggest to other organizations issues related to that, but we have to consider what's coming up to us. I don't think that we're going to get involved in actually

creating the measures that we necessarily think are helpful.

DR. BURSTIN: And just to build on that just briefly, so, the issue that has come up repeatedly when education measures come forward is it's often difficult I think particularly for patients and purchasers to be comfortable with the idea that it is a clinician checking a box that I educated a patient.

So, I think there has been very much a sense at NQF in our Content Standards Approval Committee that you're nothing about the patient without the patient and those measures have been traditionally so much more difficult to built, but that is absolutely where I think we need to go. So, just wanted to add that.

MEMBER MILLER: And there's such a variety in what constitutes patient education in diabetes. There are a hundred different components.

MEMBER BREEN: And I was just going to comment on that. One of the challenges or limitations of I think groups like this is that the measures come from the data. Right? And if there's not data out there, it doesn't mean that there's not a problem or an issue. Right?

And one of the challenges that we've had even looking at diabetes education is that some of the data is pretty lousy in terms of how the study was done, how the education was done. No standardization of the education.

So, it's not that we don't know that education is important, but if we're relying on the data to drive the measures, it's not going to happen because the data hasn't been there.

I think the data's out there, but someone has to collect that. Right?

And so, encouraging our thinking how you develop a measure without data or what

box that needs to sit in is I think really challenging.

MEMBER MCCOLLISTER-SLIPP: One question I have and the previous discussion was a great segue into it, is there was a reference to some sort of pipeline recommendation process that you just opened which I think sounds really encouraging. I have never heard of it.

But, I think I mean if there's nothing on there, I know some people who'd like to make some recommendations for things especially within the Type 1 community, those of us who wear CGMs. I mean there's a lot of interesting stuff being done about time and range or ambulatory glucose profile or whatever.

So, if there's anything that you have in terms of an announcement, I think it would be great to send it out to some of the people who are working on these measures to say hey, here's this process. Sure you've

still got some work to do in terms of the research to support it and building the evidence profile and all of that, but I think that would be something I know that the patient community would be particularly excited about.

DR. BURSTIN: Just one last point. We are actually in the process of trying to get funding to do a design session we're calling a measure incubator to allow those sort of innovative ideas to come forward. Hopefully marry them with data, funders, and experts and create those measures more rapidly, but it's sort of coming.

But, again, anything you can share on those gaps or those emerging concepts, we'd be delighted to bring it forward.

CO-CHAIR GOLDEN: And just as an aside, one of the things I noticed that it really isn't any measures separating Type 1 and Type 2 diabetes. It's all one. That's interesting. Yes.

MEMBER SHWIDE-SLAVIN: There's also the issue with education. When somebody is documented as having education, at what point did they have education?

The most recent -- American Diabetes Association just came out with new nutrition recommendations and within there is the recommendation for everyone to be sent for education when they're diagnosed. Most people are not sent for education until there's a complication and I think if there was a measure, it would help physicians to refer people to people and programs that could provide education. Not just the physician doing the education.

CO-CHAIR ROSENZWEIG: Yes, I think you're absolutely right with respect to that, but then you have to consider whether or not these particular programs are available geographically in lots of different areas and so, the different geographic areas may not be able to be judge similarly if they don't have

ADA recognized programs within the vicinity of their area and things like that.

MEMBER SHWIDE-SLAVIN: But, there are also -- there may be a registered dietitian in the area and the registered dietitian could do medical intrusion therapy which would cover education at least for the nutrition aspects of what a person needs when they have diabetes.

So, there may be something or someone, or pharmacies. There's a lot of pharmacies that are beginning to do education and so, there may be a pharmacy that's in the area and that be a recognized program.

I think that there's a lot of need to look into this.

CO-CHAIR ROSENZWEIG: Yes, but the issue has come up as to whether or not it's a -- that the individuals are certified diabetes educators or not. Is that required?

MEMBER SHWIDE-SLAVIN: These are big --

CO-CHAIR ROSENZWEIG: What's the content of the actual education? I mean these are things that have become very complicated.

MEMBER SHWIDE-SLAVIN: Yes, I know.

CO-CHAIR GOLDEN: Let's have two last comments. So, Jessie and then --

MEMBER SULLIVAN: Well, I just wanted to underscore that that last conversation takes us back to the thing of are we setting standards for what patients should receive or are we setting standards for what a physician should deliver? Because if a patient should receive it, the fact that it's not available in the area is a flaw in the health care delivery system that needs to be addressed.

But, if the standard is that a doctor should deliver it, the doctor really can't deliver it if it's not available and if the measure's holding the doctor accountable, that's -- so, I just think that's one of the

contradictions we're really dealing with.

MEMBER MCDERMOTT: I just wanted to go back to the concept of measurement developers and the pipeline. Going back to the surge of measures that we had five or six years ago, one of those measures falling off the radar out of NQF endorsement because the measure developer, what I'm hearing is, doesn't want to maintain them.

And that's a crime because if we have a good measure that's developed based on good standards, that measure should be maintained and there should be some accountability or some pick up by somebody else.

So, I'm hoping that as we accept measures and approve measures in the future that there will be a certain amount of accountability associated with those measures to maintain them. Because if they're based on drugs, LOINC codes, lab tests, whatever and even diagnosis codes for ICD-10, the fact that

they're now being dropped is really a disappointment to those of us in the industry that are trying to follow those standards and knowing that they have to be updated in order for them to be credible and then it begins to look like a variance.

So, it's just another piece of measurement development that needs to stay.

CO-CHAIR GOLDEN: And that is the challenge.

MEMBER MCDERMOTT: Yes.

CO-CHAIR GOLDEN: The process of specification is expensive and nobody is paying for it and that's a real challenge for everybody going down the road.

MEMBER MCDERMOTT: Yes.

CO-CHAIR GOLDEN: So, it's a problem.

It is a little after 10:15. We have about a ten-minute break. So, then we'll get back to do the real work I guess.

(Whereupon, the above-entitled

matter went off the record at 10:19 a.m. and resumed at 10:38 a.m.)

CO-CHAIR GOLDEN: Okay. It's about that time, and we are going to, I guess, start with doing the measures themselves and the voting and the talking. And we are going to start with Measure 59. Some of them are interrelated.

We get assigned a certain amount of time, probably because it is the first measure, this one will take a longer period of time. And we will then learn from ourselves in the process so that the other measures will go quicker. So we won't necessarily panic too quickly about time spent on the first measure.

Know all good things or all confused things come to an end at some point, and we will try to keep ourselves on task. So please don't be offended if Jamie or I say that we have to move on, or we have to focus our comments. You know, we do need to try to keep, as best we can, on some sort of a

framework and some sort of pathway to getting all the work done.

So we will do a little creative nagging here and there to keep people moving forward.

CO-CHAIR ROSENZWEIG: Yes. I think this is especially -- you know, obviously, the people who are the measures developers who are here want to be heard as part of this. But we certainly need to keep things focused. There are a tremendous number of different measures that we have to go through.

CO-CHAIR GOLDEN: So I think that what we will do -- okay. You want to start with the measure developer doing it? Okay. So I was going to say, what is the format? The format would be having someone introduce the measure in about three minutes. You're going to do all of them at the same time. Okay. All of them at the same time, since they're all yours, so that will give you a

little extra time.

And then we'd have the primary discussants give an overview for -- a very brief overview of your general impressions, and the secondary person then get into the components we have to vote on and talk about them specifically. Does that sound reasonable?

CO-CHAIR ROSENZWEIG: So you're saying NCQA is going to present all of the different measures together or --

DR. BURSTIN: The first four.

CO-CHAIR ROSENZWEIG: The first four. Okay. So that's 0059, 575, 57, and what's the fourth? 55. Okay. All right.

MEMBER KIRKMAN: So, I mean, just to comment, I realize that might be more convenient for NCQA, but I'm going to find that very confusing, if we are trying to talk about multiple measures or talk about the fourth measure two hours after hearing about it.

DR. BURSTIN: And maybe just sort of an overview of the suite of measures, because they are kind of all related, and then perhaps they could jump in before each of the individual measures if there is something you guys want to have specific. Is that okay? Okay.

CO-CHAIR GOLDEN: You know, actually, I would think -- I think 59 and 75 are related; 57 and 55 are kind of separate. Maybe just do the first two. Oh, is that to start?

DR. BARTON: Why don't I just try doing hemoglobin A1c together. And if you want me to come back, I'll be happy to speak out before you consider the next thing.

First of all, I just wanted to thank you all very much for inviting NCQA to participate, not only in today's meeting, but in the thoughtful conversations that the working groups held, which really, you know, involved a lot of very insightful

conversations and enabled us to hone our thinking and our preparation for this meeting and for thinking about the measures going forward.

In terms of the, you know, NCQA's role, I think it probably is something most of you are familiar with. NCQA is not only a measure developer, but an implementer of measures, a user of measures, and we take very seriously the maintenance of the measures that we develop, which is not to say that we don't sometimes decide to not maintain something, but it's usually for very good reason and something that we think about hard before we drop it.

The other thing that I wanted to just mention is that the -- you know, because you did see the recommendations of the MAP to the value-based measures set to CMS. There's an appeal to saying, "We don't need any of these individual components. We can just use an all or none composite measure." And I

think that there is a good reason to think that high-performing systems and highly coordinated teams can do well with all-or-nothing composite measures.

I worry that the state of U.S. health care is not consistently 100 percent in such high-performing teams, and so I think that the fact that there are also the potential for unintended consequences with all-or-nothing composite measures is also true.

NCQA has a comprehensive set of diabetes measures. You are going to be hearing about them one by one, but we view them as a set that works together and that is -- you know, that have driven -- it has driven a lot of quality improvement.

CO-CHAIR GOLDEN: And just, again, for -- never mind.

DR. BARTON: So hemoglobin A1c, we have three measures. One, 0059, is poor control, the percent of patients who have a

hemoglobin A1c greater than nine percent. 0575 is good control, the portion who have a hemoglobin A1c of less than eight percent. And then 0057, testing, measures aim to bracket good enough care.

I think, you know, the Goldilocks principle is not too high, not too low, and with -- I think it goes without saying that the testing measure is there because you need to -- if you're going to manage against something, you need to have that information. You can't manage against an A1c without having that information. So that's the reason for having the testing measure.

There is one thing I wanted to say about the physician level measures, because several of these have physician level counterparts. The reliability information that we have on the physician specification is from our recognition program, which is made up of practices that have volunteered, stepped forward, paid money to be recognized

practices.

As part of their scoring of this program, they have to achieve approximately 60 to 70 percent performance on most of the measures. And as a result of that, the beta-binomial approach that we use for determining reliability in health plans which says this measure spreads people out well, it helps me to determine good from bad, does not provide us with the same kind of information from our physician practices.

That is not to say that if it was a -- if we had data from another source, if we had, you know, excellent data from PQRS maybe, or if there were other more diverse samples of clinicians, that the measure would perform that way, but it is just -- it's the only data that we have available to us.

And so I think that the -- in the future, I'm not even sure that we would present reliability data in the same way, and I think that's something that, you know, we're

really interested in working with NQF to think about going forward. But that's just one point that I wanted to make, and I thank you for letting me have the floor.

CO-CHAIR GOLDEN: So the first measure, 59, poor control. Ingrid was the reviewer. So do you want to make some comments, some initial comments about it?

MS. TIGHE: Yes. Ingrid, I'll just jump in. We'll ask you just to give your kind of overview of the measure, and then we'll ask you to discuss the 1A evidence piece of the committee discussion that you had in your workgroup, and then we'll stop there and open it up for discussion at that point.

MEMBER DUVA: Okay. So what do you want me to start with?

CO-CHAIR GOLDEN: I think we're going to start with just a general overview, what you thought about the measure, and then we go into the components, and we'll go into the evidence, and so forth.

MEMBER DUVA: Okay. So the measure -- I think you can see it here -- has the title Comprehensive Diabetes Care: Hemoglobin A1c Poor Control, so greater than nine percent. Basically, the population is for the 18- to 75-year-old patients with diabetes, and that is Type 1 and Type 2.

The measure is to screen the patient for their A1c, and then the cut point is nine percent as a definition of poor control. So we reviewed the measure in our committee workgroup, and you can scroll down to our comments if you want to, in terms of going through the areas, the importance that I think -- this is just a quick summary, but the workgroup felt like that the -- it rated high in importance, that there was evidence to support that the --

I'm sorry, that the evidence did not necessarily support the cut point of nine percent, although evidence did support that the higher glucose that the A1c represents was

associated with poor health outcome, so that it was important from that aspect.

There was a performance gap that was noted, and that was particularly related to the physician practices and the reliability issues that the developer just mentioned. But in terms of the health plans, the reliability was strong for the measure, not for the physician practices. There was a performance gap. I just mentioned that. And, let's see, what am I missing? I have missed one of the -- was it the usability --

CO-CHAIR GOLDEN: Just to clarify, when you say "performance gap," do you mean practice variation? Or do you mean missing the standards? There's a difference.

MEMBER DUVA: Oh, I'm sorry. I'm sorry. So there was two things going on, so we looked at the validity of the measure, and then we looked at the reliability of the measure. And with the reliability of the measure there was poor -- there was good

reliability within the health plans but not with the physician practices. And we can scroll down to look at the developers --

CO-CHAIR GOLDEN: We'll get to that later.

MEMBER DUVA: Okay.

MS. JOHNSON: Maybe let's stop there and do evidence first, and then we'll have you come back and go through, because we're going to vote on each of those separately.

MEMBER DUVA: Oh, okay.

CO-CHAIR GOLDEN: But, again, a quick summary, was the general feel of it positive? Mixed? I mean, just in terms of your overall approach to the measure.

MEMBER DUVA: Oh. The overall approach to the measure, I think there was -- it was positive from the perspective that we know that patients whose diabetes is out of control will know. The evidence supports that patients whose diabetes is out of control have

poor long-term outcomes.

The concern was the cut point of the nine percent. We have had that discussion already this morning that that is not necessarily supported as strongly by the evidence. And that's it.

CO-CHAIR GOLDEN: So why don't we, then, go into the evidence discussion.

MEMBER DUVA: Yes.

MS. JOHNSON: And if I might just cut in here just for a second and draw your attention -- if you haven't already noticed, we have the evidence algorithms in front of you. So handy-dandy cheat sheets here for your algorithms.

And this is what we call an intermediate outcome measure, so we are expecting to see some kind of quantity, quality, and consistency in the body of evidence, either through guideline grading, or through summaries of those -- of the QQC. So we will help you go through the algorithm

today if you need to. That's how you will rate.

And I think we've done this exercise in the workgroup, so hopefully you're starting to get comfortable with that. But if you have questions about that, please let us know as we walk through this one especially.

MEMBER DUVA: So our overall rating of the evidence was moderate, and that was based on the fact that the evidence are supported primarily by guidelines. There were some systematic reviews that were presented by the developer, but they weren't exactly on point with this measure. So we didn't have the quantity and the quality and the consistency of evidence to consider.

However, the evidence that we did consider was supportive of the outcomes that I discussed previously, and we can go to that page where the developer presents the evidence and the gradings. You have that in your SharePoint if you want that up.

CO-CHAIR GOLDEN: And your concern about the evidence, I mean, when you say it's poor, I mean, the purpose of the measure --

MEMBER DUVA: I didn't say it was poor.

CO-CHAIR GOLDEN: Oh, I'm sorry.

MEMBER DUVA: No.

CO-CHAIR GOLDEN: Moderate.

MEMBER DUVA: Sorry. We rated it moderate, and that's based on our algorithm and the level of the evidence that was presented to support the measure. And the developer presents it in this paperwork, but it primarily comes from the guidelines, which is positive.

However, the systematic reviews were not exactly on point with the cut point of the nine percent, and the quality, quantity, and consistency of evidence was not presented. So, therefore, moderate was the highest rating --

CO-CHAIR GOLDEN: Okay.

MEMBER DUVA: -- that it can receive.

CO-CHAIR GOLDEN: Comments from the other committee members about that?

(No response.)

So let's -- I'm going to rely on --

CO-CHAIR ROSENZWEIG: I could make one comment, if I'm -- if Chair, I don't know if I -- but the original -- you know, originally when this particular measure set was created the cut point was 9.5 percent, but that turned out to represent too small a population within the overall population of patients as time went on to be able to effect an improvement with. So it eventually was changed to nine percent. And certainly there is a continuum of increased risk as you get to higher and higher A1c's.

So with respect to -- is your concern with respect to the specific amount, the specific cut point of nine percent as

opposed to nine and a half or eight and a half percent? Obviously, the higher the A1c, the greater the risks to the patient. So the evidence certainly confirms that. Is it the issue that nine percent being defined as poor control is the problem?

MEMBER DUVA: Well, we were talking specifically about the evidence, and so the evidence didn't directly address nine percent as the cut point.

CO-CHAIR ROSENZWEIG: Okay. So the evidence demonstrates the poor outcomes for the patients, and that we felt was strong, but it wasn't specific to the nine percent. That was a little bit more arbitrary. Does that answer your questions?

CO-CHAIR ROSENZWEIG: Yes.

MEMBER DUVA: Okay.

CO-CHAIR GOLDEN: So -- okay. So -- I'm sorry.

MEMBER KIRKMAN: Just, you know, my comment about the evidence is that some of

this evidence that, you know, really high A1c's are associated with really poor outcomes is very old. I mean, it's sort of -- I guess this is a philosophical comment in some ways, but it is kind of like the evidence that, you know, the higher the people's blood pressure is the more likely they are to have a stroke.

And so there is not necessarily going to be, you know, an updated, systematic review of something that has kind of been known for a long time. It is a little bit different from the evidence for, you know, a specific intervention. But a lot of the sort of observational epidemiological evidence linking poor control to poor outcomes is from, you know, the DCCT, or the Wisconsin retinopathy studies. I mean, it is very old evidence. That doesn't mean it's bad evidence, but it's not necessarily going to show up in a systematic review that has been done, you know, more recently, or at all.

CO-CHAIR GOLDEN: But, conversely,

no one would say anything over nine is good control.

MEMBER KIRKMAN: Right. But, I mean, I think -- I mean, I think it ends up being okay, because it is still going to be moderate level. But I just think that, you know, for some of these things you are -- the evidence is so embedded into the distant past and into our knowledge of everything that has come since that it is not necessarily going to come out as high level on the algorithm. If that makes sense.

MS. TIGHE: Just a process point, we can only have three microphones on at one time. So if you're not speaking, please turn your microphone off.

MEMBER TAYLOR: Is this the point where we're going to talk about the evidence? So I think the blood pressure analogy is a really important one for us to consider, because in -- we're not going to spend much time on blood pressure, but there is that

continuous graded risk relationship.

And there is also very strong trial evidence that when you move down the blood pressure curve your risk goes down. It is much more difficult, although there is UKPDS 33, and there is DCCT, and there are a couple of other things that we can use to make that argument, it is harder to tease out that kind of risk relationship that when you lower the blood sugar you, you know, universally improve outcomes.

You know, with blood pressure you reduce stroke, you reduce MI, you reduce total mortality. It is much harder to show that with the -- you know, lowering blood sugar, lowering A1c improving outcomes, especially the cardiovascular outcomes and total mortality.

And we also have those scary findings from things like ACCORD where tighter control means mortality goes up. So we probably ought to acknowledge something, I

would think, about the evidence as we go through it and say that there is a lot of reason to believe that lower glucose is better than higher glucose. But it's not powerful slam-dunk that you might have, for instance, that -- in the blood pressure.

The other point, while I have the microphone, is I'm a little concerned about the two little letters N/A for non-applicable about the unintended consequences of doing this. You know, I don't know how we fold that into our purview. Sue was kind enough to send me a reference at the beginning, because I know evidence base is everything.

But, anecdotally, there is a lot of people who are, for instance, getting their A1c's aggressively managed who don't fall within the guideline, people over 75, people with multiple risks, and so on, who suffer consequences of hypoglycemia and get hurt by this. The big unintended consequence of any guideline is the time that is taken to address

this is not taken to address other things -- the opportunity cost. And I'm not sure they are big or small, or how they fit in, but I would think it would at least take a moment of our time --

CO-CHAIR GOLDEN: Well, we'll get to that and usability and all sorts of other issues. So -- but, yes, we'll get there.

Jessie?

MEMBER SULLIVAN: I guess, Bill, I have just a question for you, because I think we are discussing the greater than nine, and it seems to me that the thing you said about the risks of hypoglycemia in some populations is an argument in favor of keeping this measure, which is looking at poor control greater than nine. You know, that the good control measures run more risk of unintended consequences, unless I'm misunderstanding what you're saying.

CO-CHAIR GOLDEN: Tracy?

MEMBER BREEN: Thank you. Just to

clarify some of the data pieces. In the DCCT trials, there was clear cardiovascular benefit on that slope of lowering blood sugar. So to be clear, on Type 1's, there is associated cardiovascular risk reduction with blood sugar lowering. And we do talk about risk reduction, and like the UKPDS trial on that slope there is strong microvascular data to support.

So I just -- for those of us who don't think about this all the time, I think it's just important to say that there has been clear data to say that on that slope there is risk reduction. I think that the issue has become how low do you go. But for this measure, we are talking way in the high end; we're talking an A1c of nine.

We can argue all day whether that's 8.5 or nine or 9.5. I think the challenge is when you look at the data, there is no data to support that particular arbitrary cutoff. But if we accept that is an

arbitrary cutoff, and how does the data around that arbitrary cutoff support it, it seems clear that there is clear risk that has been documented at greater than that number.

So it seems to me that that's a -- you know, for lack of a better number, that's a reasonable arbitrary cutoff to hang.

CO-CHAIR ROSENZWEIG: Yes. I would just echo that and say that the evidence related to poor control and the microvascular complications is indisputable, I mean, through many, many different studies. And it is only more recently that the connection between cardiovascular disease has been shown in long-standing patients.

But the issue of nine as opposed to, let's say, eight or various others, it's my understanding -- and people from NCQA could address this -- but the HEDIS, you know, you list HEDIS measures on a yearly basis, and they have steadily come down somewhat. But nine sort of tracks eight, and, I mean, the

same groups that have improvement in nine also have the same improvements in eight. I mean, there is not really a distinguishing factor between -- in any of these cutoffs, is there? Or could you just address that?

DR. BARTON: It's true that the -- there is a high correlation that -- between the less than eight and the greater than nine measure. And so what that leads us to think about is, you know, practices that are doing -- you know, paying close attention to hemoglobin A1c are hitting the mark of, you know, that sort of not too hot/not too cold, sort of Goldilocks picture that I was referring to before.

In terms of the -- I don't know that I could say -- they don't correlate perfectly, and I can't imagine that I know enough about the way that they are used in different places to say, for example, one might argue, if these were highly correlated you only need one of them. You don't need

both.

And I think that actually depending on the practice and the issues related to that patient population, and the issues related to that team, and the resources available to them, there may be some practices that are driven by one and others that are driven by the other. It's an open question, and I can't pretend to be an expert on that.

CO-CHAIR GOLDEN: Thank you. For those of you who haven't looked at the DCCT trial in the last 15 years, you remember the complication rate was not linear. It was hyperbolic. So as you get down below nine and eight, it begins to level out. So, but that would be -- so that's a part of the issue also.

Maybe we are ready to -- oh, I see one more down there.

MEMBER DUDL: Just as I would echo the need to keep both, I am the diabetes lead for Kaiser National, and I can tell you when

we try to deal with the over nines, we are dealing with much more of a behavioral issue. It is no longer, you know, technical getting information back and forth. Over eight is much different. So I do think they are different populations. I do think both measures are valid and valuable.

MEMBER MILLER: Also, remember in this denominator is included people who have not had an A1c measured at all. So that is really a big component of this, too. It is not just people who are poorly controlled. It is people whose control we are not even measuring. So just to keep that in mind as we discuss.

CO-CHAIR GOLDEN: We might be ready to vote on evidence. So is it a scale? Is it a yes/no? Tell me what --

MS. TIGHE: It's a high, moderate, low.

CO-CHAIR GOLDEN: High, moderate, and low.

MS. TIGHE: And insufficient.

MS. JOHNSON: And insufficient. So in this particular one, there is an option number 4, insufficient evidence with exception. That is not -- that would not be an option for you today, because we are not talking about exceptions. So your choices are 1, high; 2, moderate; 3, low; or 5, insufficient.

CO-CHAIR GOLDEN: So what are the implications of voting for 1, 2, or 3?

MS. JOHNSON: If you vote for 1 or 2, we will continue discussing the measure. If the -- and it used to be straight majority. It is not quite straight majority, but basically threes and fives mean we stop discussion of the measure. We don't go forward; it just dies.

CO-CHAIR GOLDEN: So a 1 or a 2 is acceptable, and anything else is not. Okay.

DR. PACE: Right. And in this case, the question 1 is -- I think someone

mentioned that there wasn't the quantity, quality, and consistency of the systematic review presented, and so according to the algorithm then that is eligible for a moderate rating.

CO-CHAIR ROSENZWEIG: I assume we are voting separately on each measure. In other words -- okay.

CO-CHAIR GOLDEN: Okay. Are we ready to vote on the A1c greater than nine, poor control, for evidence?

MS. BAL: Yes. Just give me one second. I just want to -- okay. So please don't put your number in until I have clicked the timer. And don't -- feel free to click as many times as you feel you --

CO-CHAIR GOLDEN: Well, I just want to make sure that people are ready to vote. So you get yourself ready. Anybody else? Any final comments?

(No audible response.)

Okay. All right.

MEMBER TAYLOR: And you'll review exactly what the question is that we are rating this way.

CO-CHAIR GOLDEN: It's the evidence of the measure. Is it high, moderate, low, or insufficient, to justify this measure to being -- for continued discussion and for inclusion. Correct?

MS. JOHNSON: And can everybody see the voting slides? What you're voting on is available there, and I'm a little bit nervous that the folks on this side of the room may not be able to see the screen over here. Can you guys see that well enough to --

CO-CHAIR ROSENZWEIG: We're also voting for evidence for use of the measure. It's not specifically, necessarily evidence for saying whether nine percent is poor control. Isn't that the case? Can we clarify that?

CO-CHAIR GOLDEN: I think usability is later. So I think this is just

-- is evidence over nine, poor control. Period.

DR. PACE: Yes. This is evidence about what is being measured in the measure. So it's about the numerator, evidence of poor -- the greater than nine percent.

MS. BAL: All right. You can go ahead and put your vote in now. Make sure you aim at me, not the screen.

CO-CHAIR GOLDEN: So look for a green light?

MS. BAL: Yes. And we have 20, I think.

CO-CHAIR GOLDEN: Okay. So that's 20 people. Okay. So 80 percent said 2, so we continue. What's next? What section is next?

MS. BAL: Performance gap.

CO-CHAIR GOLDEN: Performance gap.

So this is a section to say, is there either practice variation or deviation, or is everybody -- I guess the question here is, does everybody -- does every diabetic meet

this goal, so therefore, this is irrelevant? Or are there people that still need to be looked after?

So any comments from the reviewer?

MEMBER DUVA: Sorry. I was trying to pull up the exact graphs that the developer included. But in our workgroup committee, we found that there was a performance gap between plans that was high. So we still feel it was relevant.

CO-CHAIR GOLDEN: Well, the committee believes there's lot of people who have hemoglobin A1c's over nine and need attention. Any discussion? Pat? Patricia?

MEMBER McDERMOTT: Has there been testing or anyone ever looking at -- I'm thinking from the health plan perspective, and I can also say looking at it when we go to measure providers. This is requiring not only that a test was done, but that you have the result and you can review the result, and in health plans -- a health plan does not have,

we all work to try to get all the results for our members. But we don't get them all, because we haven't been able to harvest them all from all of the people that do lab testing. So we are --

CO-CHAIR GOLDEN: I think your comments are for usability. So let's hold on that, perhaps.

MEMBER McDERMOTT: Well, it's a bias that might be contributing to this variability, because you're getting more people where you just don't have the test. And, therefore, it looks like they are bad performers when, in fact, it has nothing to do with --

CO-CHAIR GOLDEN: Again, this measure may not be just for health plans; it could be for practices and for physicians and --

MEMBER McDERMOTT: Right. I'm speaking for --

CO-CHAIR GOLDEN: -- for

populations.

MEMBER McDERMOTT: I'm rooting for the provider as well. When you are trying to use administrative data to figure out whether -- how well a provider is performing, managing his diabetic patients. Without the benefit of electronic medical record or doing chart review, we have to use administrative data to know that a provider has done the right thing for his member. And there is a huge challenge in sometimes gathering all that information.

CO-CHAIR GOLDEN: I believe that will be under feasibility.

MEMBER McDERMOTT: Okay. That's great.

CO-CHAIR GOLDEN: So that will be under feasibility. So right now, the question is -- on the table is, are there -- I think the question on the table is, if every diabetic or most diabetics are under nine, then the measure is irrelevant because there is no performance gap.

DR. PACE: And, actually, it's related also to this actual performance measure. So part of the performance gap is how this performance measure identifying -- so the question is, are all health plans doing well on this performance measure?

CO-CHAIR GOLDEN: It's not just health plans. It's all providers.

DR. PACE: Right.

MEMBER KIRKMAN: So most of the data are for health plans, because most of it is HEDIS data, other than the DPRP data, or whatever. But, I mean, to me -- I mean, again, I'm thinking simplistically, but there is a gap identified because, for example, in the Medicaid health plans, you know, the proportion meeting this measure is much lower.

And so, you know, again, I don't -- I don't think of this so much as a physician measure, but it's more of a population measure or health system measure. I think somebody said, you know, these are

sort of unique patients that are -- you know, that are difficult and have lots of struggles, and so forth.

So, you know, I think there is a gap, but I don't know that it's just people delivering bad care as opposed to the system is not --

CO-CHAIR GOLDEN: But just to be clear, okay --

MEMBER KIRKMAN: -- doing well.

CO-CHAIR GOLDEN: -- this is not specified just for health plans. So, for example --

MEMBER KIRKMAN: I thought it said at the top that it was for health plans.

CO-CHAIR GOLDEN: Well, I mean, right now this measure is being used by FQHCs to assess the performance in managing a population. So I -- it can be used broadly.

MEMBER KIRKMAN: It says level of analysis, health plan, integrated delivery system --

MS. TIGHE: Yes, I'll jump in. I apologize. We were supposed to update this, but I guess we forgot to. If you look at the next -- the measure information form, which is actually what the developer submitted, it contains the correct information, that this is a clinician-level measure and also a health plan-level measure.

MEMBER KIRKMAN: Okay. But anyway, there are big differences between different systems of care. So to me, that is a gap.

MEMBER BAILEY: I'd just like to raise an important issue to address a point made by Patricia earlier. So if there's a claim available for hemoglobin A1c testing, and the value's not available, that doesn't necessarily appear in the denominator, correct? So it's only if a hemoglobin A1c level has been checked and the value is not within the target range or there's no claim, that's when it's included in the

specifications. So there wouldn't necessarily be a penalty there.

CO-CHAIR GOLDEN: I believe the denominator is anybody with a value, and the numerator would be those that had --

MEMBER BAILEY: Or no evidence of a hemoglobin A1c.

MEMBER SULLIVAN: Not for this measure. This measure is anyone with diabetes. And if they don't have a value, they fail. And if the value is greater than nine, they fail.

CO-CHAIR GOLDEN: That gets into specification issues. So --

MEMBER SULLIVAN: It does mean something for understanding what's being measured.

CO-CHAIR GOLDEN: Right.

MEMBER McDERMOTT: It can directly relate to the rate. That's what I'm getting at. So if you say you have variability in the rate, the question is, is it the member's care

and the member's stability? Or is it that you just don't have the data because you can't see it?

MEMBER BREEN: I have a question. If we're trying to define what the gap is, right, is it a gap amongst patients with diabetes, or is the gap amongst members, or is the gap amongst health systems? I think either way we define it, we're going to find that there is a gap, right? So I think if we take -- is there a gap between zip code A and B? Yes, there's a gap. Is there a gap between plan A and B? Yes, there's a gap.

So I think just to simply it, it seems that there is clearly a gap no matter which way we define it. I don't know if anyone wants to comment on that.

CO-CHAIR ROSENZWEIG: Yes. You've got the HEDIS data right there in front of you, and there certainly is a gap if you can see the numbers.

MEMBER MILLER: I was going to

comment that the gap is very wide going from the diabetes recognition programs of about 12 percent to some of the others that are about 76 percent when we are talking about, say, the 50th percentile. So we've got a tremendous gap, you know, and obviously the diabetes recognition programs are going to skew our numbers completely.

But I think if we are talking about a performance gap, I think there is a tremendous gap that exists. I also think that regarding administrative data, throughout every measure there is going to be a bit of a problem with administrative data because administrative data is never current. It always lags behind the performance of something, so that the administrative data we're given may not represent all of the things that were performed because not all of the bills have been submitted and paid yet, if that makes any sense.

CO-CHAIR GOLDEN: So, again, to --

I'm going to say that you're going to have to help me. I'm going to just keep trying to refocus us. So the discussion on the table is -- we have a measure with some evidence. Is there a performance gap, just in general with patients or with performance of the system?

After we do this vote and this discussion, we go into Criteria 2, which gets into scientific acceptability of the measure and its properties. So the discussion that came up about the numerator and the denominator is appropriate there. So there may be issues on how it is measured, but right now the issue on the table is, is there, in general, a performance gap in diabetes care with poor control or good control?

So the issues that were brought up by Patricia about -- you know, about -- the issues of how the measure is constructed will come up shortly, but not right now. Does that make sense? Maybe?

Okay. Are we ready to take a vote

about performance gap?

(No audible response.)

Seeing no cards, seeing no coughing --

DR. PACE: And I just want to make one other comment, that this is also where if there is evidence about disparities by population subgroups that that would be also considered as part of performance gap. So just for future reference.

MS. JOHNSON: And another reminder, you will be using the generic scale. So you have -- at the back of algorithm 3, this is your generic scale, which reminds you of how to think about this rating scale.

CO-CHAIR GOLDEN: So, once again, a vote for 1 or 2 continues the -- is acceptable; 3 or 4 is unacceptable. So give us a shout when you're ready for us to vote.

MS. BAL: All right. Please go ahead and vote.

CO-CHAIR GOLDEN: Okay. So, if

people are happy --

MS. TIGHE: Sorry. I'm just going to jump in so we have it in our transcript. So we have 17 votes for high and three votes for moderate.

CO-CHAIR GOLDEN: So the next item is going to be priority. Is that correct? Okay. So the next issue is 1(c), high priority or high impact, does this address a significant health problem? Prevalence, cost issues, et cetera, et cetera.

MEMBER DUVA: When we discussed this, we -- the workgroup decided that, yes, this was a high impact problem with a high cost associated for the microvascular and macrovascular outcomes that have been shown to be associated with the poor glucose control that leads to the HbA1c greater than nine. Does anybody else --

CO-CHAIR GOLDEN: Does anybody want to --

MEMBER DUVA: -- on the committee

need to comment on that?

CO-CHAIR GOLDEN: Does anybody want to question or disagree with the committee discussion?

(No response.)

Perhaps we're ready to vote on this item. So why don't you get that set up.

MS. BAL: All right. Go ahead and vote, please. And make sure you point at me.

We're still missing two people, so if everybody could just try to vote one more time to make sure we got everybody that would be great. Thank you.

So we had 100 percent, all high, 20 people.

CO-CHAIR GOLDEN: So now we're moving on. And just since -- I know I've got my cheat sheet. So just so you know what's coming up next, give you some sense for this focused discussion, the next item will be about reliability of the specifications. The next one will be validity of the

specifications. Then we'll discuss feasibility, then use and usability, and then overall recommendations. So that's the sequence we are going to be following going forward. Okay?

So now we get to reliability of the specifications and reliability testing, which means when we say about reliability, is it consistently collected, correct?

MEMBER DUVA: Okay. So here's where we run into some interpretation and probably opportunity for discussion. But in terms of the reliability -- and we commented on this -- the developer mentioned that they may not report it the same way. But in terms of the data that we got, the reliability was strong amongst the health plans, and it was not strong amongst the providers, but that data came from a -- can you say the name of that program again? Diabetes Recognition Program.

So I don't know if you want to

discuss at this time the numerator and denominator and spell out exactly --

CO-CHAIR GOLDEN: Why don't you at least describe the numerator and the denominator, so everyone knows what fails and what passes, and so forth.

MEMBER DUVA: Okay. So the numerator statement are patients whose most recent A1c level is greater than nine percent or is missing a result or for whom the A1c test was not done during the measurement year.

The outcome is the result of the A1c test indicating the poor control of diabetes, so the denominator statement would be those patients 18 to 75 years of age by the end of the measurement year who had a diagnosis of diabetes, Type 1 or Type 2, during that measurement year or the year prior to the measurement year.

So we had some discussion about that on the workgroup call, but I don't know if the developer wants to comment on the

numerator and the denominator in terms of the time period. That was one of the questions that we had on our call.

So they're 18 to 75 years by the end of the measurement year or prior to the year. Does that leave an opportunity to miss patients who are turning 18 legitimately? No? Okay.

DR. BARTON: They just have to have reached their 18th birthday by the end of the period being measured. So I think that that would not lead anybody to be missed on that end, but --

MEMBER DUVA: I mean, for the reporting of it, would you potentially miss those patients if they weren't 18 yet when you saw them, but by the end of the reporting year they were 18? Because they may have had the diabetes diagnosis that year or the year prior, is that --

DR. BARTON: It would not -- the date of the visit would not be the determining

factor. So it's your -- you've reached 18 by the end of the measurement period, and then they look back to see whether you had any qualifying diagnoses or medications in the relevant years.

CO-CHAIR GOLDEN: So it would strike me that if you have a panel of patients or an enrollment, you can identify diabetics. If you are in a fee-for-service environment, you don't know for sure the patient is still in your practice or not, I would assume, so the denominator would be difficult. Is that a fair statement? Is that discussed by your committee?

MEMBER DUVA: I mean, I feel like we didn't come to a conclusion about that in our committee. I mean, in general, it seems like a fair assessment. Maybe somebody else on the committee wants to discuss, but I can't speak for the committee to say that we felt like that was exactly spot-on.

CO-CHAIR GOLDEN: So a question

for the NQF staff in some ways. I mean, the denominator statement -- I guess it gets into feasibility and everything else, but the universality of its utility diminishes by how the denominator is defined. Is that -- how do we deal with that issue in terms of the endorsement process? And I'll get --

MEMBER LEDDY: I want to make a comment about that age range.

CO-CHAIR GOLDEN: We'll get to that in a second. Okay. Let me hold that for a second. I want to get through --

DR. PACE: So could you say a little bit more about the question about the denominator that you have? Because it's too broad, is that what you're saying?

CO-CHAIR GOLDEN: No, it's actually too narrow.

DR. PACE: Okay.

CO-CHAIR GOLDEN: What it's basically saying is you know who the diabetics are in your practice. And if they don't show

up, that counts against you in the numerator. And if you're in a fee-for-service environment, you don't know if someone has moved away, you don't know if they are part of your practice. You know, if you have a panel and you are assigned a panel, okay, you have a universe.

DR. PACE: So a couple of things to distinguish here. Under reliability, we are talking about, are the specifications such that people could implement them consistently? And then we're also looking at reliability testing results. The question you're asking, about is that going to be a valid indicator of quality that we want to get at under validity?

CO-CHAIR GOLDEN: Well, the other issue is, I could say that in terms of reliable -- or using -- being able to apply it, some people could and some people couldn't. That's the problem.

MEMBER DUVA: That's why -- I'm sorry; I didn't mean to talk so loud. That's

why I brought it up right now is I just -- I needed -- I didn't feel like I could represent our workgroup to say that we had definitely said that, yes, this was something you could reliably institute because of the denominator. And so I just needed that spoken to or the rest of the group to address it.

CO-CHAIR GOLDEN: And the technical -- so if the question is, if some people could and some people couldn't, what does that mean?

DR. PACE: So that may be more a feasibility issue in general or the usability issue. So, again, if you have these specifications, could you implement it consistently? But I think your question is, is it that every health plan couldn't do it? Or is it if it's used outside of a health plan situation? That's your main concern. And I don't know if the developer wants to respond to that question.

DR. BARTON: I think that the

development of performance measures for known denominators is years ahead of the development of performance measures for fee-for-service where you don't have a known denominator. And I would, from a parochial point of view, say that the measures that have been developed and used now over a decade in health plans are much higher bar measures and more consistent with what NQF has been espousing and encouraging us to do than really most of what you'll find in the PQRS system, because of the fact that that's -- whoever comes in your door that day, it's really not designed to enable clinicians to do planned care or managing the care.

But I don't think that that is actually a fault of the measure that we use in health plans necessarily.

CO-CHAIR GOLDEN: I don't want to -- my debate would be if you're in a practice and you have a universe of tests that you've done, you can determine of the people you've

tested how they've done, and you can't avoid that you've done the test, because you have already been judged on whether or not you did the test.

So this measures misses a universe of practice opportunities to do measurement. That's my concern.

MR. REHM: If I can just add something. Measures don't live in a vacuum. These measures are used in a variety of programs. You know, the ACL program, they're used in PQRS. Each program -- AF4Q -- everyone has their own rules of the road, their own guidelines driving this. They have their attribution requirements.

And I think if a clinician was just individually interested in their population, they would probably look for people with diabetes, either using the specification here or some hybrid, and go and see if those things were done. I don't think it's -- I think attribution is a fascinating

world, and there is a lot of competition about whose attribution rules are better and what is getting at the true thing. But I think from a spiritual level if you will, these things are doable, but there are different rules of the road. Unfortunately, you do have to --

CO-CHAIR GOLDEN: I would disagree. If you're saying attribution can shift around your denominator, that's a real problem. That is a significant problem that you just can't -- I mean, if the measure is insufficiently specified, that attribution could be all over the place and it's independent. That's a problem.

DR. BARTON: I do not think that we were saying that the specification is all over the place. When -- and guidelines may be too inside baseball a term for us to be able to explain, but I would say that the implementation of HEDIS measures in health plans relies on a set of guidelines that are things that don't even show up in these

specifications. You know, how much of the year does a patient have to have been enrolled in your health plan for you to consider them your patient? Those are the kind of things that are considered guidelines.

To me, that does not connote an "all over the place." I think that a program that uses measures has to have guidelines, and it's the responsibility of the program to create guidelines that work for that program. I would suggest that in your practice you would not just look for the universe of people who had tests.

If you wanted to hold yourself to a high bar, you would look for the universe of people who had filled hypoglycemic scripts that you wrote and look at all of those people for who had achieved the outcomes or the process measures that you set out for yourself. You would want to take the best indication that you could of, who are all of your diabetics? You probably would have set

up a registry a few years ago.

MEMBER SULLIVAN: I just wanted to speak to an experience with that. So the Westchester New York Diabetes Coalition did a project about six years ago where we took the HEDIS measure and applied it to practices, and these were not practices that tried to achieve recognition for best practices. These were community health centers, rural practices, Medicaid practices for the most part.

And the biggest change that we saw was in people who had not been tested and then became tested. That was the greatest improvement that we saw was people who had been lost to care got found.

CO-CHAIR ROSENZWEIG: In line with this question, I just wanted to ask the measure developers, I mean, the diagnosis of diabetes, which is usually listed as a secondary diagnosis, it's usually not the first one on the list, it tends to stick to someone. Once you're diagnosed with diabetes,

you know, you don't lose your diabetes.

So do you have evidence that going back two years, which is the way you specify at least with respect to the identification of diabetes, that going back two years captures the full amount of patients with diabetes, do you have any evidence related to that?

MR. REHM: Yes. You know, in the -- on our submission it's Section SA -- these are esoteric little headings, but it talks about it's the patient with at least two outpatient visits, observation visits, or non-acute inpatient encounters on different dates of service with a diagnosis and/or patients with at least one acute inpatient encounter with a diagnosis or patients with one ED visit with a patient diagnosis; or, on the pharmacy side, patients who are dispensed insulin or hypoglycemic agents during the measurement year or the year prior.

So we feel that that adequately captures, you know, the population of

interest. It's multiple things.

MEMBER DUVA: So I raised that during the reliability discussion because I just wanted the group to discuss that that can affect the ability to implement this consistently, which is what reliability is. But I also know that in the one place it says that the level of analysis is for health plan and provider, and then when it also includes private practice. But I think we are taking the private practice group out of this. It is intended for the health plan and the provider. Or is that just how you tested it for your reliability? Which came up very strong for health plan and we have already discussed that.

Not as strong for the physicians, but that had some reporting -- a lot of noise I guess is how it was defined.

MR. REHM: If I can respond. If you believe that the health plan -- and I welcome Aetna's -- or Hudson Health Plan's

perspective -- if you think of the health plan as really a distillation of provider practice out in the community, then you would say that there is a direct connection between those.

The fact that either the PQRS program, the way it's designed, is capturing its own kind of self-selected group, and that the Diabetes Recognition Program that we happen to implement captures its own self-selected group notwithstanding, that is the data we have available. Unfortunately, it's not one that shows a large range of where we can compare and contrast and we can say that's better and that's best. So it's all we have.

I think if we had nothing on the physician level, I think there are certain -- pardon the use of terms around reliability, but there's a certain face validity about that the data would extend, that measuring those patients at the provider level would be essentially a smaller version of what you are

reporting at the health plan level.

So in some ways maybe we do ourselves a disservice by having a program that reaches 3,600 physicians around the country who like to hang their hat and say, "We do a good job around diabetes care." In the same way, PQRS may be doing a disservice because there is about 30,000 or 40,000 people reporting on the diabetes measures in that program, and, again, self-selected because they have to pick some measures to report, and they picked those.

So, in many ways, I'd like to think of them as people trying to do a really good job and doing it well. The fact that there is not a lot of variability in their performance notwithstanding I don't think should spiritually undermine the measure, but I can appreciate from a raw testing perspective you'd like to see something better. But maybe wider use of these measures at the physician level where we see that data

could be instructive.

CO-CHAIR GOLDEN: Other comments on this issue?

MEMBER MAKAROFF: Just going back to this idea of how you define the denominator, so I'll just -- in the Health Center Program, our experience is that we ask health centers to be responsible for their patients, and a patient is defined as a patient who has one visit. So a patient may come to the health center for an acute visit and never come back, and that becomes part of the population.

And so I think that's an issue not just for this measure but probably a lot of measures of how we define our population, and we -- you know, is that fair? You know, as a physician, no, I don't think so actually. But, you know, it's sort of like what we have and how we look at the population and how we're managing the population and really encouraging registries and population health

management. So just a comment to add to the discussion.

CO-CHAIR GOLDEN: Other comments on this issue? Denominator? Yes. Okay.

MEMBER McDERMOTT: From a health plan perspective, the continuous enrollment is helping to control what a health plan is responsible for measuring when you look at your diabetic population. So that is kind of creating a bar for the HEDIS measures.

The issue of having to have a test is still an issue for health plans, because when we don't have a test we have to go do chart abstraction to find it in the records. So then we are dealing with samples, whereas if we had -- we were able to limit the denominator to those people where we have testing, and then say, "What's the effect?" we would not have to do administrative data polls.

From the provider perspective, when you look at the guidance from NCQA, with

the original guidance when they first published their physician-specific measures, they talk about the concept of attribution there and fairness and how to do attribution. And there are certain measures that have been developed, for example, by the AMA that say you can look at this member to see if they have CHF, and if they have this drug, based on two visits within the year. And they specify physician attribution.

The HEDIS measures do not, but we take -- we have done research on the concept of one visit. And if a member has something based on that one visit, we give the doctor credit, else we look for a second visit, and often within a longer period of time, to make sure that they have seen the patient more than once and it's not just a single visit for a sore throat before we assign attribution.

That happens to be how we do it within Aetna. I believe that Cigna, Unita, all the others, have come up with ways to --

and have looked at their data and have talked to their physician population.

ACOs are -- we know who that population is, and we are doing a metric supporting them, just to give a flavor of what's going on in the industry based on getting these kinds of specifications and figuring how to use them to get valid information.

MEMBER DUCWORTH: Okay. The denominator inclusion criteria, they become essentially constraints for people like me. I can only use certain metrics and certain programs as indicators or assays.

Now, when we are looking at this particular metric, I think if we are -- the testing, that population that doesn't have that A1c, that can be good for P for P programs as a carrot or an extra stick for providers who aren't, one, performing that A1c.

But, James, I agree with you, it

doesn't necessarily give us an accurate representation of the population that we are evaluating. So I think if we are going to go forward with this type of -- or with this criteria in the denominator, then I think it is kind of our responsibility to really be specific to organizations on how we use this particular metric.

It should not be assumed that this represents the health of a population or a panel or necessarily the -- maybe more so the performance of the physician. It leans more towards the outcome or, I'm sorry, of process, because you're combining two approaches.

CO-CHAIR GOLDEN: Other comments on this issue?

(No response.)

Somebody -- you wanted to talk about the age range. The age range might belong in usability or feasibility. I don't know. Does it belong here? Validity. Age range belongs in validity. So it's not in

reliability, but it will be in validity. I continue to have you in the parking lot. I'm sorry.

MEMBER DUVA: So the next thing we're discussing is validity.

CO-CHAIR GOLDEN: They have to vote so --

MEMBER DUVA: Oh, we have a vote?

CO-CHAIR GOLDEN: Karen?

DR. PACE: I just want to make a comment about the -- and this is probably more the validity issue, but it has come up several times about the measure construction, and that if you don't have a test result it goes against you in the numerator.

This is one suggested way from our Consensus Standards Approval Committee of constructing a measure so that, you know, the issue is if -- if the patient has -- you know, if you can't find the lab results, maybe that's a problem as well. So it may lead you to a different kind of improvement efforts if

you discover that the reason for your bad score is because you don't have lab results. That may have a different solution than if you're really having patients with greater than nine percent.

But it does combine things in a way that drives to overall improvement, and it is one suggested way of constructing measures.

CO-CHAIR GOLDEN: Has the NQF done anything about standards about attribution or any kind of consistency?

DR. PACE: No. And that is an ongoing issue, right.

DR. BURSTIN: Get through SES and risk adjustment, which is the big one at the moment, and then we'll work on that one. We'd like to.

MEMBER McDERMOTT: Just one other point, if I could make it, is that there is a diabetic screening hemoglobin A1c measure separate from this measure that is consistently done by the HEDIS and in provider

performance. So you already know the diabetics that are never getting screened. Just a thought.

So this is adding on that population that never gets screened, plus those that have a level greater than nine. Yes, they have to be a diabetic. Right. But then there is another measure that is simply saying how many diabetics have not had annual screening.

MEMBER KIRKMAN: Annual testing, I guess is -- I thought you meant screening people for diabetes.

MEMBER McDERMOTT: Hemoglobin A1c. Hemoglobin A1c testing.

CO-CHAIR GOLDEN: I don't see anybody looking for attention here. Are we ready to vote? Yes. So let's vote, and it's reliability.

MS. TIGHE: Okay. Go ahead and vote. Everyone keep pushing until we get to 20. Sorry. If everyone could try again.

Okay. There we go.

All right. We have five for high, 13 for moderate, and two for low. So we'll move forward.

CO-CHAIR GOLDEN: Okay. So now we move up to -- excuse me, I've got my cheat sheet missing. Validity. There it is -- validity. Thank you. And the concept here is, does it actually test what you want it to test?

DR. PACE: Right. And this actually includes quite a lot. It is -- you know, are the specifications consistent with the evidence that was presented? And then formal validity testing, or I think in the case of this measure face validity was what was done. But also, what we can term "threats to validity," which has to do with, you know, who is excluded; for outcome measures risk adjustment, are there actually meaningful differences in performance; if there are multiple specifications, do you get comparable

results; and that -- so it's a combination of all of those things.

MEMBER DUVA: Right. So in our workgroup we discussed that this measure did have strong face validity in terms of expert consensus in the ability of the measure. This measure also lined up well with other measures of quality for diabetes, which supported the validity.

So from that perspective, the measure had high validity in that also we spent a lot of time discussing the threats I guess to validity, which would be patient factors that cannot be controlled for, concerns about -- there was a small discussion about stratification because of different population groups that -- where the gap was higher in the different health plans versus Medicaid/Medicare I believe it was.

So there was some concern about that, and then of course there is the discussion we just had about whether or not

it's specified correctly.

CO-CHAIR GOLDEN: Okay. Comments on this one?

(No response.)

Are we ready to move on to a vote?

MS. BAL: Go ahead and vote, please.

MEMBER KIRKMAN: We've voting very specifically on different categories, but we are just being -- it flashes up there and we are supposed to vote, and we have to turn this way to vote. So I don't know if anybody else is having a problem with this, but I just wish somebody could read what we are voting on. I mean, I know we're voting on --

DR. PACE: Right. So --

MEMBER KIRKMAN: -- validity, but, I mean, the specific --

DR. PACE: Right. And just a couple of things that, first of all, under validity a measure can only get a high rating if there was empirical validity testing of the

performance score. So this measure is relying on face validity, so you would be talking about a moderate reading at the highest level, and then you would go from there.

So, but -- so, you know, this is where you're considering, you know, will this be a valid reflection of quality of care? And some of the things that you look at here is, you know, how it is specified, who is excluded, are they the right exclusions, you know, does it actually distinguish -- you know, indicate meaningful differences in performance across those being measured.

And so, you know, if it's an outcome -- I know this is an intermediate outcome that is not risk adjusted. I don't know if you had discussions about that or discussions with the developer about that. But for outcome measures that might be a consideration under validity as well.

So it is, you know, taking all of that into account, you know, in general to

give it a rating, and it needs to get a high or a moderate to continue.

MS. JOHNSON: And let me just put in here they actually did do some empirical validity testing. They did some correlation analysis. So --

DR. PACE: Okay. So -- sorry, I missed that. And so that it's eligible for a high rating.

CO-CHAIR GOLDEN: I have a question before we vote. I guess it was a question for our colleague from the -- from HRSA, for Laura. Periodically, I have people from FQHCs say that you need to risk adjust for socioeconomic status. Has that been -- I was just curious, in general, how HRSA views that kind of commentary? I don't know where that sits and whether it's valid or not. I just --

MEMBER MAKAROFF: Yes. It's -- the question of whether it's valid or not, I don't know that I know the answer to that, but

that's something we hear a lot, too, especially for health centers that service special populations, which we define that as serve a high percentage of homeless populations, migrant seasonal farm workers, things like that, as well as, you know, generally speaking I would probably say the health center population all has socioeconomic factors, you know, that influence their care and their outcomes.

So as far as we're -- what we do about that, so we ask health centers to report their actual performance on our measure set. We have like 12 or 14 measures that we collect annually, and then we have an adjusted quartile ranking methodology that we go through that adjusts for some of those things. So we kind of compare health centers to other like health centers.

So it adjusts for things like percentage of homeless, percentage of uninsured patients that a health center may

care for, things like that, to be able to see kind of relative performance that way. But I'm happy to talk with you more about it. I'd love your insight, too, or anyone else's. I think it's something that we spend a lot of time sort of thinking about and how do we -- is it worth adjusting for?

I mean, this conversation actually happened yesterday in my office, too. It was like, you know, HEDIS doesn't, to my knowledge, adjust for, you know, other socioeconomic factors. So I don't know that any other programs are doing that, but that is something that we kind of do to look at relative performance.

CO-CHAIR GOLDEN: Has NCQA ever discussed this or looked into this issue?

DR. BARTON: NCQA has. NCQA is against adjusting away socioeconomic differences from a belief that there is no reason why we should expect -- seeing evidence that excellent care can be provided to

challenging populations, to then excuse away that responsibility is not consistent with the overall mission of improving health care quality.

And I think actually what our -- the HRSA representative just described is actually stratification, comparing peers to like peers, which is different than adjustment, which tries to make everybody comparable to each other, you know, using statistical techniques. And so I think that I just wanted to make that distinction between stratification and adjustment.

MR. REHM: And if I can just tag on to Mary's comment, we do have two measures in the HEDIS set that are risk adjusted. Those are our plan all-calls readmission measure, because we perceive that as an outcome measure and necessarily needs that adjustment at the health plan level of specification, and then also our relative resource use measures. Those are five

measures that are looking at resource use cost and quality. So we felt that that's appropriate as well.

DR. BARTON: But they were adjusted by health conditions, not --

MR. REHM: Correct.

DR. BARTON: -- socioeconomic.

MR. REHM: Not SES. Right.

CO-CHAIR ROSENZWEIG: I should mention, though, I mean, in the data that you have all showed us with respect to each of the plans, I mean, in almost all of the categories the Medicaid patients did worse than the -- than the HMO plans.

Now, does that mean that they're -- the Medicaid patients got worse care, or does it mean that there was an adverse selection? You don't know.

DR. BURSTIN: And I'll just mention that NQF is in the middle of doing a pretty significant body of work on this very question of SES and risk adjustment with a

draft report out next month. This month. Sometime in March. So we really welcome your input on this. This has become, obviously, an increasingly high profile issue as more and more measures are being used for higher stakes uses, including patients selecting providers as well as payments. So more on that to follow, but it's obviously an important issue.

MEMBER KIRKMAN: I think one thing is that I don't think you can just narrowly look at a measure like this as a measure of the quality of care, like, you know, one physician with one patient, because these are generally patients that just -- there is just lots of issues going on.

I mean, I think more broadly you can think of it as a measure of how our entire system doesn't do well with particular kinds of patients or patients with particular, you know, socioeconomic or comorbidity, psychiatric comorbidity, things like that, but I would hesitate to say that this is by itself

just a measure of quality of care, at least on the kind of micro level, because I think it's -- I think you're going to end up sort of beating up a physician or a health care system for things that they probably can't really control. But if you look at it as sort of our whole system, or lack of system of care, then maybe it is.

MEMBER SULLIVAN: Yes. At Helen's invitation, I did want to comment on what Mary said. I think, to me, one of the really important things here is the difference between stratification and risk adjustment. If you risk adjust, you don't know what's going on.

So we know that black women have poor birth outcomes every time it's tested. So if you risk adjust, that goes away. You don't know that. But if you don't stratify, you can't figure out who is doing better within that population because the only -- the biggest correlation is between race and

outcome, and that is all you see, if that's what you -- so I think the way HRSA does it is the right way. You don't risk adjust, but then you stratify.

CO-CHAIR GOLDEN: All right. Are we ready to vote?

MS. BAL: Go ahead and vote, please. So we are still just missing a few more. If people could just try to make sure that we're getting everybody's results. Thank you. Perfect. Thank you. The final results are high, seven; moderate, 13.

CO-CHAIR GOLDEN: Okay. Feasibility. So feasibility is -- again, make sure we have our concepts down -- extent to which the specifications, measure logs require data that are readily available but could be captured without undue burden and to be implemented for performance measurement.

MEMBER DUVA: Despite -- I know the challenges of the administrative data, perhaps we didn't consider that enough in the

workgroup, but we rated feasibility high, and this is a measure that has been in place and it is being currently reported.

CO-CHAIR GOLDEN: Comments? Jessie, are you up, or are you -- okay. We have kind of been discussing this for the last little bit anyway. Any other comments?

(No response.)

Ready to vote? All right. The --

MS. BAL: Please begin.

CO-CHAIR GOLDEN: -- polls are open.

(Laughter.)

Okay.

DR. BURSTIN: Data collection can be implemented.

MS. BAL: So we have high, 14; moderate, five; and low, one.

CO-CHAIR GOLDEN: All right. Use and usability. Correct? So this is the extent to which potential audiences -- consumers, purchasers, providers, policymakers

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

So a quick question on that one for definitions. You know, I haven't been around too long. The use of the word "and" versus "or" -- accountability or performance improvement, accountability and performance improvement, there's a big difference. So can you elaborate on that for me?

DR. BURSTIN: As I mentioned earlier, at this point it is an "and." But I think one of the questions is, is there recognition that, you know, going forward there may be some measures that are potentially suitable for one versus the other? And do the criteria need to change with that? So broadly we are asking you about the measure for a wide range of potential uses.

CO-CHAIR GOLDEN: Comments on this issue? Usability. Okay. Bill?

MEMBER TAYLOR: Is this the point at which we bring in the issue of unintended consequences? So building on what Anna said before, you know, if I have a hemoglobin A1c above nine, is this -- you know, do I have -- does the physician then have higher performance standards if he makes the -- her or his practice inhospitable to that patient? If the patient has mental illness or English is not their first language? Are there things that would happen as a consequence of this standard where care might be -- instead of improved, it might be degraded unintentionally. Is this an unintended consequence of making a standard like this? Is this the place where that --

CO-CHAIR GOLDEN: I have seen practices who were not -- have non-adherent diabetes tell the patients to leave the practice, so it becomes --

MEMBER TAYLOR: Yes. That's a good example. So but this is -- we're voting

here -- is on this topic. It includes this notion of unintended consequences. Is that correct?

DR. PACE: Right. But also, you know, it helps if there is some evidence about that versus the theoretical or anecdotal stories. And to look at that in weighing in relationship to the benefits, so you want to weigh both the benefits and the potential unintended consequence.

CO-CHAIR ROSENZWEIG: Yes. It's my perception -- and maybe -- I'd be interested in hearing from other people -- that the issue of cherry-picking has always been raised with respect to this kind of situation. But, in fact, there is not a lot of evidence that -- certainly in large groups that such actually occurs.

So it's -- but it is more anecdotal than anything else. But if people have other evidence to present, that would be of interest.

The other issue, of course, that is always raised by endocrinologists is that endocrinologists would be caring for mainly patients who have high A1c's that are referred to them. So if they're compared with the primary care doctors, there may be problems in terms of evaluating those kinds of things.

MEMBER BREEN: And this comes back to stratification again. You know, do you compare hospital-based clinics to hospital-based clinics that have a very different patient population than their faculty private practice two blocks down the road? I think it's the same issue we have already discussed.

But, again, I don't think there is any expectation that any measure should have zero harm, right? We're talking about a balance in benefits versus harm. And even though that they're -- I agree, I don't think there is any data out there to suggest that this cherry-picking process is going on. I

don't know that anyone has looked to see if this cherry-picking process is going on. So lack of data doesn't mean that the concept is nothing.

MEMBER KIRKMAN: So two things. One is there is evidence from the UK where they put in a very aggressive pay for performance system that there actually was very little cherry-picking. Of course, that's a very different system from ours. But I don't know whether this is the time to bring up whether the -- whether performance is improving over time, because my understanding with this measure it has remained pretty steady.

And I think -- I suspect that's partly because I'm not sure that this really measures quality of care so much as kind of bigger issues that, you know, are -- as a society we are not able to fix very well. But I just wanted to throw that out there, that this -- I mean, I actually really like this

measure, and I think it's really important. But I don't think we are seeing changes in the proportions, unless I'm reading the data wrong. It has remained pretty fixed, the proportion of patients that are above nine percent. Is that right? Or am I wrong?

DR. BARTON: I'm not sure if I understood exactly your point. But I would just say that, you know, the median is one thing, and then another question is how the 10th and 90th percentiles are going. And I think there is no question that there are places that are improving through the -- you know, the issues that you were discussing earlier, that this is a particular set of patients who you have to go after with different tools to actually get them into care. And so there are places that have been very successful at doing that.

So I think, has the whole nation moved? I'm not sure that it has. But have there been pockets of improvement driven by

attention to this? I would say yes.

MEMBER DUDL: Yes. Let me just respond to that. We just got through interviewing the top 10 performers in A1c's over eight and nine, and what was very interesting is all of them actually do more of a population base where they go after looking at all of the people. We all know that there is a top 10 percent, that you are not going to move 65 visits and nothing happens.

But it turned out there were quite a few that were in the panel but just not coming in. So I don't think this is exhaustive.

MEMBER MILLER: I think when I wrote some comments to myself about this measure and use and usability, regarding the question of if performance is improving for glucose control, I made a note that it appears that each type of health plan had results that have varied within a small range up and down over the course of years.

And so, you know, I really thought that a lot of that had to do with, you know, practice level management skills. But it also may be reflective of what you were discussing earlier with a roaming denominator. That may also account for some of the small variations that we are seeing up and down year to year.

CO-CHAIR GOLDEN: It was my impression that the FQHCs had seen some improvement. Is that -- are other folks looking at their data?

MEMBER MAKAROFF: With this measure in particular there has been like -- from what I know, which is annual measurements, we have one data point once a year for all 1,200 health centers. That number hasn't really changed in the past three years since we've been measuring.

MEMBER BAILEY: Just to address the question of changes over time based on anything that was published in The New England Journal of Medicine last year, '99 to 2002, 18

percent had hemoglobin A1c's greater than nine; 2003 through 2006, 13 percent; and 2007 through 2010, 12.6. So there is positive movement towards lowering A1c. So it is a select population, but, still, evidence on a nationwide sample, that there is positive movement there.

MEMBER KIRKMAN: Although for some reason the NHANES data are always different than the HEDIS data, right? I don't know why. Because in NHANES they are actually measuring A1c's on a selected, you know, representative sample.

CO-CHAIR ROSENZWEIG: I do think -- my recollection is that, yes, 10 years ago there was steady improvement in the HEDIS data as well, but it seems to have flattened out in the last -- at least certainly in the last three years.

CO-CHAIR GOLDEN: So, again, what are we voting on? So make Sue happy here. Accountability, transparency, everyone can

read it perhaps.

DR. PACE: Right. Usability and use includes, you know, is it being used and can be used in accountability and transparency programs. So public reporting, pay for performance, accreditation, et cetera. And the expectation is that they -- you know, especially on endorsement maintenance, that the measure is being used.

4(b) is about improvement, because the whole point of endorsing these is to make improvement. And then the third element is about the unintended consequences. That has been raised. So taken together overall.

MS. BAL: Okay. Please vote now. So the final results are high, nine; moderate, 11.

CO-CHAIR GOLDEN: So we're in the home stretch. Overall recommendations for endorsement. So it's a yes or no. Probably don't need a lot of discussion on this one, given how we have been going. Does anybody

want to have further discussion on this measure for overall endorsement?

MEMBER KIRKMAN: Can I just ask a question? Because, I mean, I just want to get back to this question of physician level reporting on this measure versus health plan reporting, because to date it has primarily been health plan reporting, other than -- and even in the physician recognition program, it is just that you pass this Chinese menu of options, so we don't really know that Dr. X, you know, is here, and Dr. Y is that.

So, I mean, so -- but is there a plan? I mean, can you see in the future where -- because I would look really bad. You know, I have a lot of people referred to me with really high A1c's that I don't necessarily get down. I mean, and I don't know whether that's really a question at this point, because I think on all of the measures we sort of decided this measure is okay, but --

CO-CHAIR GOLDEN: That's going to

be the second hour, the Brandy conversation I think.

MEMBER KIRKMAN: Okay. And maybe it should have --

CO-CHAIR GOLDEN: No, you have a good point, and that's --

MEMBER KIRKMAN: Maybe it should have come up in the usability.

CO-CHAIR GOLDEN: Yes.

MEMBER KIRKMAN: It does concern me if this, in this future, is going to be publicly reported on the physician level, because I think it could have a lot of sort of unintended -- it is kind of an unintended consequence. You're sort of punishing people for something that is really not a quality of their care.

CO-CHAIR GOLDEN: Jamie and I had a side bar, and that, you know, one of the problems -- we talked about attribution, coming back to your attribution missing link, is that if you're in practice and you get

attribution done by the payer, you often get the patients attributed to you at the time of measurement so you don't know you're being measured on the patient, which gives you little time to react.

So, yes, there are some issues. So --

CO-CHAIR ROSENZWEIG: There are other issues that have come up, especially the idea -- actually, this occurred even more when there was a less than seven measure. But plans may be accountable, but what they then do will then institute a variety of procedures to make their individual providers accountable, such as pay for performance, or tiering, or a variety of other things.

So even though HEDIS might just hold the plans accountable, it does filter down to the physicians as an unintended consequence in many cases.

CO-CHAIR GOLDEN: Do you have a comment?

MEMBER DUVA: Well, just I know we vote now whether or not we recommend the measure. When is it that we talk about parsimony and -- that Karen mentioned earlier, you know, in terms of all the measures when you're looking across the board at all the measures and if there is redundancy or some that are better than others. Do we do that at the very end?

CO-CHAIR ROSENZWEIG: Ready to vote? Not yet?

MS. BAL: Go ahead and vote. So the final results are yes, 20.

DR. BURSTIN: So it always takes an hour and a half for the first measure. I just thought I'd put that out there. Never seen it happen in any less time. You'll speed up. Don't worry.

CO-CHAIR ROSENZWEIG: Of course I thought the conversation was good, and I think it sets up some of the other discussions for later. So that's a big help.

So next, 575. So A1c's under eight. Okay. They're calling in?

MS. TIGHE: Yes. We have some folks -- I'm sorry. So we do have some folks who just call in right at the appointed public comment times, and so we are trying to hold true to them. It's awkward timing, since we only got through one measure.

But since we're pretty close to 12:15, I do just want to pause and see if we have any NQF member and public comment either on the phone or in the room. Yes. It's a commenting free-for-all for those who are looking to comment. You can provide comment on whatever you would like.

MR. LEE: Thanks so much. I am David Lee, the Executive Director of the National Bone Health Alliance, which is a public-private partnership on bone health that includes 51 organizations from public specialty society and nonprofit sectors as well as industry as well as four government

liaisons. And we are here to -- I guess I'm here to talk about the three osteoporosis measures that will be looked at this afternoon, which are very important to our constituency because they really support fracture prevention programs which have not been widely utilized here in the United States, other than closed systems like Kaiser and Geisinger.

And I think especially the exciting part, one, because they will help address the narrow 80 percent post-fracture care gap. It is also because I know that our hope is that if they were to be endorsed by NQF today, and through the process, that they would become a potential new core measure set that the Joint Commission would use in terms of reaccreditation, which I think is a very important stick for folks, because if you see the kind of flat-lined, you know, care gap that really has not changed much in a long time, and certainly lack of awareness, both by

health care professionals and consumers about post-fracture care and osteoporosis, I just want to make sure that we emphasize as -- our full partnership are fully behind this and fully prepared to engage with our health care professionals and consumers that we can reach to help make this a reality and to really change the face of osteoporosis care here in the United States.

Thanks so much.

MS. TIGHE: Operator, if you could see if anyone on the phone would like to provide a comment at this time?

OPERATOR: If you would like to make a comment, please press star and then the number one on your telephone keypad.

Okay. At this time, there are no comments.

MS. TIGHE: Thanks. Apologies for that untimely interruption. Turn it back to you.

CO-CHAIR GOLDEN: So we --

technically, we're going to have lunch sometime soon. On the other hand, we could continue moving, so -- I'm sorry? 12:30 is the lunch? Okay. So let's get moving along, then. That's fine.

So who was assigned this one? Sorry. Vicky, okay.

MEMBER SHWIDE-SLAVIN: So I agree. My fingers are crossed that we can move through this one quickly. It's just like 59, except it's at the other end. We're looking at A1c control less than 8.0 percent. This is for patients 18 to 75 years of age with Diabetes Type 1 and 2, whose most recent level was below -- oh, you still can't hear me? Sorry. Was below 8.0. And develop a rationale -- is that the measure is critically important from both a clinical and financial perspective because the largest improvement in outcomes occurs by a reduction of blood sugar levels in those patients with the highest glycohemoglobin level.

One second, my -- okay. So shall we just jump right into evidence? Okay. So our group felt like that there was sufficient evidence to support this measure, and I think we only had one real comment where someone did express concern on whether or not this measure was good in general to evaluate a population, but we didn't go into intense discussion around that. Do you recall that, Ingrid?

MEMBER DUVA: The discussions about the evidence, it demonstrated increase mortality, and I think it was -- Bill, do you want to comment on this? The patient's -- the increased mortality of patients when their A1c gets too -- gets in the tighter control group?

The ACCORD study, there is a lot of references to the ACCORD study because that was --

MEMBER TAYLOR: Yes. I mean, I think ACCORD is the one where with tighter control total mortality went up, which was a big red flag to people. And it -- you know,

for old people, you know, the first time there was a big diabetes control study and outcomes were measured goes back to UGDP. I'm looking down at the end of the table for people who know about that, you know, but that's back in the 1960s where the first question about increased mortality came up with tighter control. That was sort of, you know, pushed aside.

But, similarly, in the UKPDS 33, there was a subgroup of the overweight people where there was a question of increased cardiovascular mortality, so the question keeps sort of percolating through the studies that -- are we doing some harm at the same time that we're accomplishing -- there's no question that the microvascular qualifications go down, retinopathy goes down, nephropathy goes down. I mean, that happens over and over again. But interspersed in the studies are these worries about either total mortality or cardiovascular events that pop up here and

there, not with great consistency, but enough to, for some people, raise concerns. And certainly with the tightest control in ACCORD, when the total mortality went up in the more aggressively treated group, it got at least a few people's attention.

MEMBER DUVA: So I think the summary of the discussion in our workgroup was the concern about the tighter control leading to the adverse outcomes, and so then the definition of the eight percent and whether or not that was loose enough to account for variability in the glucose readings reflected by an A1c of 8.0.

MEMBER KIRKMAN: So just to -- I know we don't want to get into a huge, long discussion about ACCORD, but just remember that the control group in ACCORD had an A1c target of 7.0 to 7.9, and they had lower mortality. So it would be a little bit hard to say that a target of less than eight percent is going to increase mortality. I

mean, we could get into the issue of the lower limit, but -- you know, and the other thing is that -- and everybody knows this, but the people that did poorly in ACCORD were people actually in the intensive arm who had the highest A1c's, not the lowest. So it's a very complicated issue, but I think this measure to me is sort of like the control group in ACCORD in terms of the goal.

CO-CHAIR ROSENZWEIG: Something's blinking. Okay. I was going to make the exact same points about ACCORD. And you can -- and ADVANCE and VADT similarly. The big issue also is that these patients were of an older age and also had coexistent cardiovascular disease.

Now, that raises the issue as to whether or not within the spectrum of patients -- I mean, the American Diabetes Association has raised its range of goals to go from -- anywhere from 6.5 or seven percent up to eight, eight and a half percent with respect

to different individuals. So elderly patients with a short potential life span and patients with multiple complications might very well manage okay with A1c's between eight and eight and a half or something like that.

The question is, what percentage of the total population is going to be affected when you're looking at broad numbers of people? I think the big issue is whether or not the supplies to the Medicare population, over 65, whether or not those patients actually form a significant part of the population.

Anyone else want to address that?

MEMBER TAYLOR: So I want to make clear what -- my position is not any kind of disagreement with people getting their A1c's below eight and restricting it to people 18 to 75. My concern is the unintended consequence that people don't always read the fine print of what the American Diabetes Association says or what these guidelines are.

And as we push clinicians by how we pay them and what the standards are to believe that tighter control is better and they should be worrying about it, the unintended consequence that I worry about -- and I don't have data to support it -- are people being treated inappropriately, more aggressively?

CO-CHAIR GOLDEN: I think it has to be for a different segment of the discussion if that's okay.

MEMBER TAYLOR: Yes. That's all.

CO-CHAIR GOLDEN: All right.

MEMBER McCOLLISTER-SLIPP: And, again, I want to raise the possibility of the unintended consequences, too, because I don't think we can discuss these kinds of things without really thinking through that. I mean, again, I've had Type 1 for 28 years. My blood sugar is very stress responsive, even when I don't feel stressed.

You know, my former endo is now

Chief Science and Medicine Officer at ADA, and there were lots of times while I was under his care that my A1c was above eight. It doesn't mean that he wasn't providing excellent care. I couldn't get any better. It doesn't mean I wasn't doing everything that I should be doing. I was. It's just difficult.

And it's hard enough to find an endocrinologist to start with, especially if you've got complex disease and you're, you know, difficult to control. I mean, I really haven't found one since Bob left clinical practice. So I think we really need to think about that because these things have a way of getting calcified.

And even though -- I mean, the ADA -- and Sue can certainly speak to this better than me, but the ADA has gone out of their way to make sure that their guidelines reflect the individuality of each patient. And the ability of each patient, given their own set of circumstances, disease progression, et

cetera, to be able to meet certain criteria that, you know, they've gone out of their way to say these are our targets, but you've got to take it on a case-by-case basis.

When we're studying these kinds of standards, I completely understand the need for them, I completely understand why we're doing this. That's why I volunteered. You've got to start somewhere.

But it's difficult to make a binary decision one way or the other when there will be real consequences of this. I mean, if the measure is widely -- widely adopted and it's used, you know, maybe some physicians will get paid more, some will get paid less. And over time you create a disincentive -- an additional disincentive because there are already lots of disincentives for people to go into endocrinology and for people to specialize in those of us who are difficult to treat.

MEMBER BREEN: So I think it just

-- oh, I'm sorry.

MEMBER McCOLLISTER-SLIPP: I was going to say the challenge becomes you are creating a tool, but we don't necessarily have oversight over how that tool is going to be employed. Right? The measure, I think we would agree, is valid, that for the majority of people with diabetes, an A1c less than eight represents reasonable or a goal of control. The question is how that measure will be utilized by either plans, is it going to be utilized as -- at an age level.

Do you carve out -- I mean, we're talking in our health system right now, who do you carve out of that, right? Who do you identify as your high-risk subpops that you don't put into that? And I think -- I don't know how much pre-thinking or advanced downstream thinking we can do on this other than to note that it's a concern, and keep bringing it up as a concern.

CO-CHAIR GOLDEN: And, again, it's

for other elements of the discussion. So I guess you could argue -- I mean, I'm just trying to make sure we're not glossing over it, but less than eight is better potentially than not less than eight. But then you have some populations issues, and so forth. So I guess the scientific validity would be, are there populations where it's not appropriate? And that would be a fair game for this discussion. So that would be like exclusions, and so forth, down the road.

DR. PACE: And the other thing is I think this is one reading, the most recent reading, I mean, the other discussion along that line is, you know, I think it was brought up before, time in range or an average, you know, so that can be brought up in the other discussion as well.

MEMBER KIRKMAN: So I think it sort of gets into a bigger philosophical -- more philosophical question, but I don't think it just applies to this performance measure,

and that is that the -- you know, more and more we realize that care has to be individualized and we have to take into account patient preferences and all of these different factors.

And so at the bedside your definition of quality of care is really going to depend on that patient. But for performance measures it has to be something that can sort of be collected simply, and it does end up sort of being kind of like a one size fits all.

So I think it's just that tension between performance measures, which can't -- you can't go into every single chart and say, "Well, this seemed reasonably good quality." You have to sort of set some limits. And I think that tension is just going to keep growing as care becomes more individualized, so --

CO-CHAIR GOLDEN: As an aside, down the road, I mean, you know, we have drawn

episodes of care and total cost management. We did it for perinatal. And we decided you couldn't risk adjust a pregnancy on a sickle cell patient, or you couldn't risk adjust a pregnant -- so we just excluded things.

So, you know, it got to the point where we say, "Look, if we covered 85, 90 percent of the pregnancies, that's not a bad deal." So, you know, it gets into the same issue. There are just some risk categories that aren't worth, you know, covering because it's just -- you can slice and dice it that well.

MEMBER McCOLLISTER-SLIPP: But since we're not in control of how these measures are used, I think it's our responsibility to think this through. And I'll just throw this out as a quick example, just because it literally happened last night after I got in from a long day and a very late flight.

I recently switched insurance

companies. I take Aranesp and Erythropoietin because I have CKD-related anemia, and you have to get recertified every three months or whatever. So I received a letter last night from my insurance company saying that the anemia drug was not medically necessary because my hemoglobin was 11.1. If it were 11 or 10.9, I would have been fine. So basically they are denying care based on this guideline which is based on studies that were looking for something very different.

So by drawing this arbitrary line, you know, it gives people the ability to deny care in a way that is completely inappropriate because it is taken completely out of context.

CO-CHAIR GOLDEN: Jessie?

MEMBER SULLIVAN: Yes. I was just going to speak to the point that had been raised earlier that the measures that we have -- when the measures are a threshold measure, we're left with this. But it's not impossible to measure improvement at the patient level,

you know, hemoglobin A1c change, delta. We just don't have measures that do that.

CO-CHAIR GOLDEN: So I'm going to push us along here and say that I think that we have some issues here we can discuss on other parts of this discussion. But I think that we -- I think we have enough discussion here to discuss the evidence for less than eight. Unless people want to violently disagree, I'll -- I will move us along. So we can I think take a vote on the evidence question -- low, high, medium, and something else.

MS. BAL: Okay. So everyone can start voting now. Let's just try one more time to get those last two in there.

Okay. So the final results are high, nine; moderate, eight; low, three.

CO-CHAIR GOLDEN: Okay. Performance gap. So are there disparities in care? Are there practice variations? Is there variation in care?

MEMBER DUCWORTH: Yes. The workgroup found that -- or we believe that there were care disparities across plans, and that there is a substantial gap, even for DRPs that have demonstrated better outcomes than health plan data, and that this measure should be possibly indicated as disparity-sensitive.

Yes, that's it. There's a disparity or a gap.

CO-CHAIR GOLDEN: Do we have people who want to address this? Sue, you're up, but I don't know if you mean to be up. There's one other there. Patricia? No. Okay. Going once. People are getting hungry. Good. So we vote.

MS. BAL: Go ahead and vote. Okay. We have high, 16; moderate, four.

CO-CHAIR GOLDEN: Okay. Now we go to impact.

MEMBER DUCWORTH: Okay. The group also feels that we do think that this is a high priority or demonstrates an opportunity

-- high impact opportunity. And rationale -- there are significant implications for both morbidity and mortality cost of care across a very large patient population.

CO-CHAIR GOLDEN: Comments? Questions?

(No response.)

Are we ready to vote? Going once, twice. Time to vote.

MS. BAL: Okay. And then just to repeat the question, the question is, does this measure address a significant health problem? And so you can begin voting now.

Okay. The results are high, 16; moderate, three; low, one.

CO-CHAIR GOLDEN: Okay. Reliability of the specifications. And just for the record, why don't you tell us or remind everybody what the specifications are. Did I confuse you? I'm sorry. I've got them here.

MEMBER DUCWORTH: For the

numerator and denominator?

CO-CHAIR GOLDEN: Yes, please.

MEMBER DUCWORTH: Okay. Patients whose most recent A1c is less than 8.0 during the measurement year, the outcome is the result of an A1c test. Denominator is patients 18 to 75 years of age by the end of the measurement year who had a diagnosis of diabetes, Type 1 or Type 2, during the measurement year or the year prior to that measurement year.

CO-CHAIR GOLDEN: What is similar to the high measure but similar to --

MEMBER DUCWORTH: Yes. And there are exclusions in this population as well. However, not that one exclusion that we were kind of upset with in 59. Exclusions include patients who did not have a diagnosis of diabetes in any setting in the measurement year or year prior. Also, patients who meet the following criteria -- a diagnosis of polycystic ovaries in any setting or a

diagnosis of gestational or steroid-induced diabetes in any setting.

CO-CHAIR GOLDEN: Now, in this case, just to be clear on the numerator, if no test was done, it would still count against you, correct?

MEMBER DUCWORTH: Yes.

CO-CHAIR GOLDEN: Okay.

MEMBER DUCWORTH: Okay. So the group -- scrolling down, excuse me. We do -- or the group did feel that the health plan data has sufficient reliability. Now, there was concern that the physician level data had weak reliability. The differences in performance by individual -- or individual providers would be less reliability distinguished, and overall moderate, high with health plans, low with providers.

CO-CHAIR GOLDEN: A similar conversation.

MEMBER DUCWORTH: Yes.

CO-CHAIR GOLDEN: Comments or

questions on this item?

MEMBER KIRKMAN: Can I just ask NCQA to explain again why the reliability is low for physicians? I know it's from the DPRD data. Is it just because of the way the data are submitted? Because this is one where I do think it's going to be provider level. You know, I think there's a lot of provider level reporting of this.

DR. BARTON: Right. So there is a menu from which practices can choose. However, what we look -- what we noticed when we looked at the physician level data, it looks like the huge majority of practices select the hemoglobin A1c less than -- not greater than nine and less than eight as measures that they want to report on.

So the mean is high and the distribution is real close, because they are all doing pretty darn well on this, and that is kind of why they want to be in the program and think themselves worthy of being in the

program. So that kind of data distribution is mathematically designed to do poorly on a beta-binomial assessment, because the beta-binomial is asking, do you have enough spread on this to be able to distinguish good performers from bad performers?

And, unfortunately, this data source doesn't allow us to access information about that at all. It is really --

MEMBER KIRKMAN: Okay. So by definition they are sort of all pretty good performers, and there is not much --

DR. BARTON: Variation.

MEMBER KIRKMAN: There's not a way you can tell --

DR. BARTON: There's not good spread.

MEMBER KIRKMAN: Okay.

DR. BARTON: Yes.

DR. PACE: Did you provide any information about sample size? Because that is also a factor when you get to the sample

size. So I was just curious with your provider level data if that was a factor.

MR. REHM: I think the sample size is 30 patients. Again, this is physician. They do this in their own offices, and it's a sequence of physicians. They start the date and then whoever comes in with that diagnosis is the person tested, so that it prevents gaming.

CO-CHAIR GOLDEN: So the numerator -- the denominator creation is much different than for the plan.

MR. REHM: Correct. Again, and then it would be different for PQRS, or it would be different for an -- I mean, it's --

CO-CHAIR GOLDEN: Yes. But that -- I mean, that gets back to holding. If the denominator can vary, then your specifications are all over the place.

MR. REHM: You know, I think the feeling is is that there are programs out there that design their programs. And CMS

designs program X way, and we are not measuring the program, we are just -- we can articulate the program that we update it from, which is our own.

CO-CHAIR GOLDEN: But, you know, I've heard Peggy O'Cain comment about HEDIS light and not happy. Then again, from the NQF standpoint, the specification of the specifications, and if you do it differently it's not -- I had a conversation with a plan person. He wanted to collect the -- we were going to use the NYU algorithm for emergency room care.

And she goes, "Oh, we'll report the emergency room use. We just will report it differently." And I said, "Well, you're not, then, reporting the measure. And so you either report the measure as specified or you don't." So that -- you know, that's a problematic issue that you bring up that you can make up your own denominator and say you're still reporting the measure. And

that's something we have to face as a group; the denominator is the denominator.

DR. PACE: So you're saying that how this is applied to physicians is it's a -- they do a 30-patient sample, so that's the difference?

DR. BARTON: That's what they're required to do for program --

CO-CHAIR GOLDEN: It is a different -- fundamentally, then, a different measure.

MEMBER KIRKMAN: Yes. But they're not using this measure per se, right? I mean, it's really --

CO-CHAIR GOLDEN: They have reported the reliability on that --

MEMBER KIRKMAN: -- because the denominator --

CO-CHAIR GOLDEN: -- on a different measure.

MEMBER KIRKMAN: Well, yes. Right. But that is why the reliability is

probably not good. And, again, it's a voluntary program, and so people with poor quality are probably not going to --

CO-CHAIR GOLDEN: No. But I would say if they've submitted reliability data on this measure using their physician reporting, they are reporting reliability on a different measure.

MEMBER KIRKMAN: But they also report it on the HEDIS data. So that's not the only reliability data that --

CO-CHAIR GOLDEN: No, I agree.

CO-CHAIR ROSENZWEIG: I'm unclear. Is the physician reporting -- that's an entirely different process. Is that still considered within the purview of this measure?

DR. BARTON: The measure -- NQF asked us to say how we were looking for this measure to be endorsed. We have this measure specified for health plan reporting, and it is used by many health plans and has been for a long time, and using these specifications. We

also have a physician level specification. It has been picked up by a variety of programs, and, like it or not, they each have their own -- they are only going to apply it to the people who are in their program.

So the -- I'm afraid I'm not understanding --

CO-CHAIR GOLDEN: That was not --

DR. BARTON: I'm not understanding your concern.

CO-CHAIR GOLDEN: If you have changed the construction of the denominator, which is what you've said you've done, it is a different measure.

CO-CHAIR GOLDEN: If you have a set of specifications, and you apply it to these people and it gives you this group, and I apply those same specifications to this other group, it gives me a different group. So the denominators are different. That's not to say that they are -- the specifications are different. So I don't --

CO-CHAIR GOLDEN: That's the whole point of specifications. Yes, Patricia?

MEMBER McDERMOTT: There's many ways that you can collect a measure. You can do it using administrative data, claims, encounters, and that's the only way you get the measure. Or you can do it through e-measures where they use their electronic medical record in order to say, "These are the diabetics." And you go into the record and you say, "Do they have this lab test?"

That's -- it's using the same specifications, and you'll find when you look at the different ways -- when you go through the documents that have been released by NCQA, whether you use the physician specifications or the health plan specifications, you can use it all through administrative data, or you can do what's called a hybrid measure for health plans, where abstractors go out to doctors' offices and they find all of these things using a random sample method.

There's a random sample method where you go in and go to the doctor's office. You're going to get higher, more credible rates than you get using the same specification, because there is more information there that is not available administratively. But it's the same specification.

DR. PACE: But this measure requires looking at the -- getting the actual value. That's not going to be in administrative claims.

MEMBER McDERMOTT: If we have -- with LOINC codes, yes, we have the administrative data. Yes.

CO-CHAIR ROSENZWEIG: You're talking about the physician recognition program for diabetes? Is that what you're referring to with this particular physician-based measure?

MR. REHM: I know this is kind of hard to grapple with. We have a program that

uses the -- what we would reference as the parent measure, which is -- has a longer history, which is the HEDIS health plan measure. And the physician groups got together and said what would be appropriate physician measurement for people who want to be recognized, and they developed that program.

We are not asking NQF to endorse that program. We happen to have data on the use of a physician level measure, and we are sharing that with you. That is --

CO-CHAIR GOLDEN: But let me just make sure I understand. So in the denominator for that program, it is 30 consecutive diabetics in your practice?

MR. REHM: Yes. That is a fundamentally different denominator. I think the intention -- you know, measure implementation, whether it's at a health plan level or a provider level or a group level, or you can cut it in a variety of different ways,

is probably going to have implementation requirements that are unique to themselves.

I think if you look at the portfolio of NQF measures, all 8- or 900 of them, I would imagine that you would not have each measure -- 600, sorry, Helen. You wouldn't have a measure that was specified perfectly for each type of use that would be out there. It would be almost impossible to --

CO-CHAIR GOLDEN: And that is part of the NQF's problem.

MR. REHM: Well, you know, I think we can -- we would love to address the totality issue. And we feel your pain because we also try to get the balance right. And I guess what I'm just trying to share with you, had we not had a physician recognition program, then we would have had the PQRS program that is using the measure.

And it's out there and it has its own rules of the road. And is it finding the

same variation in practice? Yes, about the same. I mean, it's -- I just think we can, you know, chew on the bone of how it's used in this program or that program or another program. It might end up being less filling, and I can understand why it's unsatisfying to have a specification that does not fit perfectly into every single implementation or use.

And I'm just trying to educate folks about --

CO-CHAIR GOLDEN: I don't want to pound this on the ground, but that's why we have -- there has been a cycle. We have rooms like this, we have rooms at NCQA, and then we have other rooms with the same people talking about harmonization of measures. And what you've just described is why nothing is harmonized, because everyone is making up their own variations to fit their programs.

MR. SAUNDERS: Can I clarify? I think the difference here, though, is that the

-- how you get into the denominator is consistent across every implementation here. These are the things that you have to trigger to be a diabetic. What is different in the implementation here, and what we're describing about the testing results, is about the sampling strategy and how you -- the validity of the performance rate. It's a different question than the denominator. The specification of how we sample cases and the data that you are reviewing is not part of the specification, and that is I think where the distinction is.

DR. BURSTIN: This is a really complex issue and we recognize this, and this is certainly bigger than NCQA, and I think it affects all of us. But, again, this measure is not about its use in the physician recognition program. It is the physician level measure below -- clinician level measure I hope below the overall health plan level measure, which at least gives us some

alignment.

So I think the alignment issue here is actually important. But, again, it is not -- you are not approving the measure that is in use in the sampling strategy for the physician recognition program. That is NCQA's program. This is the measure specifications at this other level of performance.

So, I mean, I know there are questions being raised because these -- the testing you have been presented is on the physician recognition program, but I think they are slightly different questions.

MEMBER DUVA: I just wanted to suggest that we stick to what the issue is, and this data was presented for reliability of the measure. Reliability is different than -- well, like any measure, you have to test for reliability in whatever population you put it in. So this measure may not be reliable in that population. We don't know because of the noise, so we don't know either way.

I think what the developer did was present to us what data they had in terms of reliability, and we cannot say that it is reliable in that specific population. But it does not mean that the measure is not reliable in other populations, and it doesn't mean that the specifications are inappropriate or have changed. It is whether or not it is reliable in that specific population that they happen to produce at the provider level, and that's it. That's all we have, and we can't read any more into it because then we're going way beyond what we have been presented.

MEMBER KIRKMAN: What she said. And I think this is a -- this is a meta issue and not specific to this measure or to NCQA or to diabetes measures. And I suggest we move on.

CO-CHAIR GOLDEN: That's fine. We should vote, unless people want to have final comments.

MS. BAL: So to repeat the

question, do the results demonstrate significant reliability to the -- sorry, I can't read from there. So the differences in performance can be identified both -- for both plans and individual physicians. And the voting is open.

We are just waiting on one more, so if everybody could just retry just in case. Okay. Perfect. Thank you. So we have high, three; moderate, 14; low, three.

CO-CHAIR GOLDEN: We go now to validity.

MEMBER DUCWORTH: Okay. The group -- here again, we rated this moderate to high. There were parallel concerns that we saw on the reliability testing that one of the comments was, yes, correlation coefficients were generally strong to very strong, that we are seeing higher validity and correlation for health plan data than physician level data. So, here again, moderate to high.

CO-CHAIR GOLDEN: Comments?

Questions?

(No response.)

Ready to vote?

MS. BAL: All right. To repeat the questions at hand, do the results demonstrate significant validity so that conclusions about quality can be made? Do you agree that a score from the measure, as specified, is an indicator of quality? Is testing adequate for both plan/system level and physician/group level? And it is now open.

So the final results are high, four; moderate, 12; low, four.

CO-CHAIR GOLDEN: Okay. Feasibility.

MEMBER DUCWORTH: The workgroup -- we agreed that the data for the measures or the data is routinely generated and used through care delivery, and that moves the EHRs and claims data, make collection analysis of this metric relatively easy, straightforward.

CO-CHAIR GOLDEN: Comments or questions on this one?

(No response.)

All right. Ready to vote?

MS. BAL: Okay. So feasibility, we're looking for data that is generated during care, electronic sources, and data collection can be implemented. And so the voting is now open.

Let's try one more time. We are missing one person. There we go. So it's high, 17; moderate, three.

CO-CHAIR GOLDEN: Use and usability.

MEMBER DUCWORTH: The group noted that the developer listed five current uses of the measure, including public reporting. There was some concern about patient factors regarding glucose control that are beyond the control of the provider. However, overall the workgroup did agree that the measure is a useful measure that is easy to use.

CO-CHAIR GOLDEN: Any other comments? This is your adverse consequences, and so forth. Didn't know if anybody wanted to say anything. I'm just giving -- opening the door for them. I cut them off before.

MEMBER KIRKMAN: So, I mean, again, I don't know whether this is the place to bring this up, but there is this issue of the last hemoglobin A1c, so, I mean, I guess it could work out either way. But, you know, you could have somebody that, you know, had been 12 and a half and is now eight and a half, and, you know, you made a huge benefit. But it just so happens that their last one is above this cut point. I think that's the whole issue with these threshold-based measures, though.

MEMBER LEE: So I just wanted to bring back up the issue of the patient voice. Of all the measures today, I feel like this one is most out of control of the physician and most in the control of the diabetic.

Diabetes care has changed drastically since I've been practicing the last 10 or 15 years, in that we have moved from much more -- less of a prescriptive way of dealing with diabetes to much more shared decision-making and having the patient have a voice in what they do.

And so I would definitely be in favor of seeing this measure perhaps reexamined or modified to include more of what the ADA has recommended, because one size is very difficult to fit all.

CO-CHAIR GOLDEN: What does the ADA recommend? I'm sorry.

MEMBER LEE: Well, the ADA recommendations that we brought before for different populations. But I think of all of the measures this is the one where the patient really comes into play, where, you know, we'll say, you know, get your -- you know, take more insulin or check more often, but it's really in the hands of the patient.

And I think shared decision-making

and patient voice should be something in consideration to a measure. That has been very valid, has been around for a long time, but I think diabetes care is changing over time.

MEMBER McCOLLISTER-SLIPP: Just to defend the patients, I mean, oftentimes statements like that -- and I don't think you meant this -- are construed as patients are completely unadherent and noncompliant. I mean, it's just -- we really don't have very good treatments at this point. So part of it is biology. Part of it is choice and lifestyle. And part of it is limitations like, you know, working three jobs and can't make it to the gym.

So just to state that for the record. It's not always an issue of choice; it's an issue of biology or circumstances.

MEMBER LEE: I apologize for that. So --

CO-CHAIR GOLDEN: Ingrid?

MEMBER DUVA: I have a question. This might -- this is clarification that might -- for later as we go through these meetings. But, Sue, what you said about the patient that goes from 12.5 to 8.5, so if you are reporting on both of these measures at once, are you still then penalized? Because either they have dropped out of your greater than 9.0 poor control group, but you still -- is it still kind of a -- I mean, help me understand, is that still kind of a penalty, then? Because now they're not within your under eight group? I just think it might come up later when we're talking about parsimony and what measures are -- you know, what is more useful to measure.

MEMBER KIRKMAN: Yes. I mean, I think in some ways it is going to balance out. But, you know, it -- I do think it's -- you know it's a little bit -- it's sort of like saying it's not good quality of care because this percentage of your people were above eight percent, or they didn't fall in the less

than eight percent, when, you know, some of them could have been moving there. But, you know, I think it's going to probably balance out in the end.

I just wanted to say, I mean, I agree completely with what you're saying about individualization and the patient voice. But there actually are -- you know, there are very few sort of groups of patients that the ADA recommends a goal that is higher than less than eight. So it's -- you know, it's the very frail elderly, with limited life expectancy, but otherwise the sort of general ADA recommendations, it is kind of six and a half to eight.

So, you know, it's -- there are not going to be huge populations of people, but I totally agree. I mean, I think it, again, gets to this -- it's difficult to do patient-centered care and speak to individualized care when we sort of by definition in a performance measure have to

say we're going to slice the pie here. I think it's difficult.

MEMBER McCOLLISTER-SLIPP: Yes. And I'm not suggesting that we should all shoot for, you know, whatever makes sense and whatever feels good and whatever is convenient. I mean, I obviously shoot to get my A1c under eight, and I think anybody would.

But I don't want to penalize my physician, and that's kind of what we're looking at. I mean, conceivably this could be used as a mechanism for limiting access to drugs. I don't think it's analogous at this point in time, but maybe it could be something like what I experienced with my anemia drug.

But I think the biggest issue is, are we disincentivizing physicians for taking care of the most complex, difficult-to-treat patients? And, you know, there is a lot of stuff that we've learned about diabetes, and we have gotten much better treatments, but we still don't have very good treatments and it's

very imprecise and it takes a lot of work. And, you know, it's easier for some people than others, and we don't really know why.

And I just don't want to discourage more people from taking on patients like me and the many others like me by, you know -- and, I mean, that's why I like -- what I like about what you guys did at ADA and what they are continuing to do, is it is kind of a target. But use your sense, use your judgment with the discretion. If we're studying a line in the sand, there will be implications for that, and I don't want -- we have already got a shortage of endocrinologists. I don't want that to be one of the implications, because that ultimately will hurt patients and it won't help the system.

CO-CHAIR GOLDEN: As an aside, I think in England in their quality programs you're allowed to exclude some people. And, you know, there are some people -- I have a couple of patients that, regardless of what I

do, they're going to not be in the ballpark on a parameter. But I don't want them to become orphans either, so -- so maybe that's down the road something to consider as well. You know, people are very brittle, and that kind of thing.

MEMBER LEDDY: I just need to speak for the practicing endocrinologists. I mean, we all know that diabetes is a hugely labor-intensive endeavor, diabetes management. And it would be good if we could put some advisory node, that there are these tough patients that are very, very hard to manage, and it would be good not to penalize physicians who care for them.

I practiced in a multi-specialty group for a number of years, and there is no question as these guidelines became more specific and limits were placed that I accumulated the toughest patients, the ones that were the hardest to manage. And, you know, happily or not, a lot of them didn't

come back because I was too tough on them. But it is a huge burden, and we mustn't forget about it.

MEMBER HAYDON-GREATTING: So this is the epidemiology view. We need a severity of illness to add to the perspective, so that we can stratify those patients. And I don't know -- we can't do it now because we don't have all of the -- we can't get it there. We have to have a massive database.

But if you can stratify patients by the severity of illness, then you can have -- you don't orphan, I mean, those patients that are out there. I just worked three years to get someone from a 12 down to a nine, and we are celebrating that. And so with this, that would leave -- my physician would be lifted out on that, but I want to honor her for busting her butt with us together to get that patient down there. And I think we all have one of those patients or 10 of those patients, or, in some cases, in Mississippi,

150 of those patients.

CO-CHAIR GOLDEN: Yes. Oh, go ahead, please.

DR. PACE: I just want to, you know, ask -- you know, I know you're having to deal with the measure as it's specified, but that's exactly the question I was going to ask is about the possibility of adjusting for severity. You know, what would be the factor that would be used? Or if there were specific patients to exclude versus just self-identifying, I want to exclude these patients. Are there specific parameters that, you know, would be supported in the evidence that should be excluded?

So just maybe for future discussion when you get to the future.

CO-CHAIR ROSENZWEIG: My question about this measure is, to what -- what does it add to the previous measure that we have just considered, which we approved unanimously? The big issue I suppose is that people would

feel that just by adopting the other measure it would encourage physicians to basically have mediocre control of their patients, that once they got them under nine percent they would just not bother to get better control.

I'm not sure that's really the case, but I think that's probably the -- that's the only, really, issue that -- to me that is -- that comes up here that we have to think about.

CO-CHAIR GOLDEN: Are we ready? I think, Anne, you've got to put yourself down there.

Any final comments?

(No response.)

Ready to vote? And we are voting on -- somebody read what we're voting on. Hopefully, we won't be that close.

MS. BAL: Okay.

CO-CHAIR GOLDEN: About usability -- it's used for transparency, used for improvement, benefits outweigh evidence of

unintended negative consequences.

MS. BAL: You can begin voting now.

Okay. The final result is high, seven; moderate, eight; low, four.

CO-CHAIR GOLDEN: And now we get to the big picture. Endorse, yes or no? Final comments?

(No response.)

Ready to vote.

MS. BAL: You may vote now.

We're waiting for two more. Let's try one more time. Okay. We have yes, 17; no, two.

CO-CHAIR GOLDEN: So they are -- we're done. It's time for lunch I think. It's time for lunch. Let's get lunch. When do we reconvene? Thirty minutes? 1:30? All right.

(Whereupon, the above-entitled matter went off the record at 1:04 p.m. and resumed at 1:42 p.m.)

A-F-T-E-R-N-O-O-N  S-E-S-S-I-O-N

(1:42 p.m.)

CO-CHAIR ROSENZWEIG: Okay. I think we'll get started again. We have a big agenda ahead of us.

I'd like to thank Bill for doing all of the heavy lifting. I think hopefully things will go smoother from here on in, and we'll be able to move a little bit more quickly.

And so, I think the next one on our agenda is going to be Number 57. And I'm looking for the list of people. Do you have the list of people who are reviewers?

MS. TIGHE: Yes. Anna McCollister-Slipp is the primary, and Bill Taylor is secondary, for 57.

CO-CHAIR ROSENZWEIG: Are we supposed to hear from the measure developers first, or have you pretty much given us -- okay.

MEMBER McCOLLISTER-SLIPP: Hello

there. I'll try to do this without getting curry on me, but we'll see.

So, I mean, I can walk through this in great detail if we'd like. I would think that this might be one that would be -- since it's a process measure especially, that it might be one that would be relatively straightforward.

I mean, we didn't have a lot of discussion around it in terms of a measure for process during our call. And, colleagues, please tell me if I'm misremembering something other than the philosophical discussions that we have already had previously about A1c as a measure in and of itself. But, I mean, certainly my conclusion of the necessity for it was that it is certainly necessary. There is strong evidence correlating A1c -- showing A1c as an important measure of control.

You know, maybe it's -- I mean, I think in my comments I described it as a necessary perhaps, but not for sufficient

measure of quality. So, you know, is it a high priority? Yes. There seems to be evidence to suggest a significant health problem, that we need, you know, to get better control of in a population. This is a great way of assessing general control and quality.

And there wasn't much discussion around that. Bill, do you remember it any differently?

CO-CHAIR ROSENZWEIG: Any other comments?

MS. JOHNSON: So maybe we can just open it up to talk about the evidence a little bit. The measure is about doing the test for A1c, so is there --

CO-CHAIR ROSENZWEIG: I think, obviously, there probably isn't much disagreement about the need for doing the test. The question of course comes up always, the frequency of the test. Why once a year, let's say, as opposed to once every six months or once every two years, or so forth. Any

comments about that?

(No response.)

Okay. I mean, obviously -- sorry?

(Off-mic comment.)

CO-CHAIR ROSENZWEIG: All right. Well, in this case the -- but the numerator statement is -- for this case is once a year. You've got to realize a lot of people aren't seen more than once a year in certain situations. And the other issue of course is that in order to be able to get the outcomes measures for the previous two performance measures that we just evaluated, you need at least once-a-year measurement on a yearly basis for that purpose.

I mean, there are -- certainly people wouldn't disagree with the fact that patients on insulin and with Type 2 diabetes -- Type 1 diabetes might need a more frequency A1c measurement. But at least once a year certainly is a reasonable number.

Okay. So any other comments?

Yes.

MEMBER DUDL: I'm wondering if -- and this is something for the NCQA, et cetera. I'm wondering if we get above 95th percentile, instead of saying, you know, one plan got to 98 and the other one got 97, you know, there is a small part of the population who, you know, you may not want to test, if they have Alzheimer's, et cetera.

I'm wondering if we couldn't say that above a certain threshold, that you've attained success and we are no longer going to rate you, because I see some people -- what happens is some people drive to get the last few percent that really aren't super important to get, instead of working on control of getting the over nines or over eights on down. It's just a thought.

CO-CHAIR ROSENZWEIG: Well, actually, it's a reasonable thought, because we are going to be getting to the performance gap issue in a minute. I think that's

actually going to be part of the discussion there.

Any other -- yes.

MEMBER SULLIVAN: I don't know if this is the right place to raise it, but I'd be interested in knowing what the developers thought about making this into a composite measure of the other processes of diabetes care. I know in our health plan, that's how we use it. That's how the state of New York is using it.

I think it has become kind of common and you get away from the high performance, then, because, you know, having done all of the measures. So I just wondered what the developers want to say about that.

MR. REHM: The humor here is that we did have an NQF-endorsed composite. It is simply not used in the market. And when you don't have a use -- thinking about use and usability -- so we withdrew it because the market said thank you, but not all that

interested.

There are clearly competing measures. Later on, in Phase 2, you'll look into the all-or-nothing measure, which is in many ways a hybrid of many of the measures you're looking at here and then some other preventive health measures.

So, you know, I think Mary spoke to why we see some advantage of having these individual measures in play. And when we talk about performance, you know, I can just point out that there is a gap, you know, a gap that we feel about 10 to 15 percent, between the 10th percentile and the 90th percentile of performers.

So when we see that, that's cautionary. When we see gaps between commercial and Medicaid, that's cautionary. And then we don't share it with you, but we have regional data that shows, as you can imagine, although it's different for each measure, different gaps in care there.

CO-CHAIR ROSENZWEIG: Any other comments related to evidence? Yes, Bill.

MEMBER TAYLOR: So we did discuss briefly, in our subcommittee conversation, that there is no evidence for the periodicity with which the testing gets done, and there never will be. Nobody has ever done a trial comparing every six months and every 12 or 15 or nine or something. So it's sort of what is reasonable and what is the sense, you know, if you do it too much, and what are the consequences, or not enough, and it's going to be a sort of face validity judgment call around the table, I think, about the periodicity question, as it is elsewhere when it is, you know, why do we do these things once a year?

The other one -- I think the -- this was one of the ones that measure applications partnership did not endorse. Is that correct? Is there something we should learn from that?

DR. PACE: Well, one of the things, I mean, it's probably related to the fact that this is what we often refer to as a distal process. And as someone pointed out, it is necessary but not sufficient. And when you look at the evidence that is presented, it is based on expert opinion. So that's something that you all will have to think about, whether you want to make an exception to our evidence criterion for this measure, meaning that, you know, most of the evidence is about the control or the treatment versus, you know, taking the assessment once a year or several times a year.

So, you know, and this was actually embedded in those control measures. So patients -- if the assessment wasn't done, it didn't -- you know, it marked against the performance. So I think that's the main motivation from the map standpoint is that it's one of these process measures. It's very distal. It's necessary, but it's not

sufficient. And they're trying to drive towards, you know, more intermediate outcomes, outcomes, and the actual treatment kinds of process measures.

Helen, anything else?

CO-CHAIR ROSENZWEIG: Sue?

MEMBER KIRKMAN: So, I don't know if we're still on evidence because it seems like we've -- we're having some non-evidence. But it was kind of my point, too. I mean, it's almost like this is already embedded in several other measures that are going to continue. And so, what's the point? I mean, maybe that's just a bigger question.

And, you know, it's a little bit like, you know, there's a lot of outcome evidence for high blood pressure or lowering blood pressure, and so, you know, if 99 percent of people are getting their blood pressure measured, do you really need a performance measure on it? I mean, it's not quite as high here, but performance is pretty

high.

CO-CHAIR ROSENZWEIG: Yes. I think that is going to be -- we're going to deal with that in the second issue related to the performance gap, I think, which is going to obviously come up.

MEMBER KIRKMAN: So I guess we should, you know, finish up the discussion of evidence and vote on that.

CO-CHAIR ROSENZWEIG: Yes. So any other additional comments about the issue of evidence?

DR. PACE: So if you -- just to remind you about the algorithm for evidence, Algorithm 1, when we have a measure that's based primarily on expert opinion, you know, Box 10, the first question is, are there or could there be performance measures of a related health outcome or evidence-based intermediate clinical outcome or process?

And if the answer is yes, then the algorithm says no exception. And then if the

answer to that is no, is there evidence of a systematic assessment of expert opinion? And you answered that. And then the last question is, does the Steering Committee agree that it is okay or beneficial to hold providers accountable for performance in the absence of empirical evidence?

So I guess the question here is, you know, first of all, kind of working through this, whether you -- you know, what we'll be asking you is to -- you would need to vote whether it meets our exception criteria to move this measure forward.

So maybe we should have a discussion about that first, the algorithm. Mary, do you want to --

CO-CHAIR ROSENZWEIG: Yes. Tracy?

MEMBER BREEN: Sorry. All of a sudden I had brain fog. We are just talking about A1c testing as a value of measuring for diabetes control, right? Is that what we're saying?

CO-CHAIR ROSENZWEIG: Yes.

MEMBER BREEN: So it's almost like we should have done this one first, right? Because this is the basic measure, like, we need to do this test in order to measure diabetes control. And then whether it's greater than nine or less than eight is really kind of drilled down on that, so I think that's maybe where my brain fog is coming in. So if we just separate that out and say, are we saying that there's data to support A1c testing and outcomes? I think there is very clear data on that.

DR. PACE: So this is, you know, where it gets a little fuzzy is that there is -- the evidence is about the control, right? And obviously, in order to manage it, you have to do the test.

And the question is, do you have to have a -- I don't think there is anyone at all that questions that it's important and necessary to do the test. The question here

is, do you need a performance measure on doing the test? Or can you construct performance measures that are based on the outcome and the treatment? That's the major question that you are addressing.

So, you know, we have a process for you to accept -- you know, to pass this on an exception to the evidence if you think it's, you know, important to continue this as a performance measure on evidence. And then you will get to the performance gap, as you have been talking about, so that may be another issue where you have some concerns. But the first question is about the evidence.

CO-CHAIR ROSENZWEIG: Any other comments? Yes. Bob?

MEMBER BAILEY: Well, I guess the major question would be here because you're not dependent on having a laboratory value, that your measurement population is larger both in terms of the numerator and the denominator. And does that provide any

different insights as opposed to having -- limiting it to the population where you have the specific laboratory values?

CO-CHAIR ROSENZWEIG: Were you going to comment?

DR. BARTON: If I might.

CO-CHAIR ROSENZWEIG: Sure.

DR. BARTON: So I would say to the initial formulation, I agree that it's a judgment call, whether you say that the evidence is only for the management and doesn't include the step of the testing, but -- so be that as it may, I think the point that you just made is absolutely true.

If the hemoglobin A1c testing can be reported on an entire population by use of administrative claims, it potentially is being reported on a much larger population of diabetics than the ones whose -- you know, there was some discussion before about the hybrid reporting. And, you know, in our -- both in our Medicaid and our commercial groups

of health plans, 95 percent of the plans do go look at charts to get the hemoglobin results for the less than eight and greater than nine measures.

So those samples are 411. That's what NCQA has determined is statistically reliable for our health plan reporting. And so, there would be a difference in the use of the hemoglobin A1c testing. I would also say that, again, you know, would that we all were at Kaiser and had, you know, 98 percent on so many things. But health care in this country is not uniformly at that level, and so I think that finding that measures look too low bar in some tables, and that if you were to go somewhere else you would find that that's what they're just struggling with, the first steps.

CO-CHAIR ROSENZWEIG: I think the big issue here is, 10 years ago, this obviously was an important measure. Is it still an important measure now?

Should we vote on the evidence

issue? Oh, comment. Sorry. Bill?

MEMBER CURRY: So from a practicing clinician's point, this process measure is embedded in the two previous measures that we looked at. But when I get the data from my payers, or I as the quality person in our medical group provide this information to my partners, we need to have a list of the individuals whom we're serving who have a gap. And if we get that information, if we get -- if we try to get that information from the previous two measures from our insurers, we'll just know if they're in range or they're over nine. But I won't know who has not had the A1c. It's there. It's part of the --

CO-CHAIR ROSENZWEIG: It's part of the numerator. Yes.

MEMBER CURRY: But we'll have difficulty culling that out. So for the provider at the field, this I think is an important piece to help them identify those

patients in their population who have the gap. I think it will be much easier for this to be information used at the provider level.

CO-CHAIR ROSENZWEIG: Okay. Any other comments?

(No audible response.)

All right. Why don't we vote on the evidence.

MS. BAL: Go ahead and vote.

CO-CHAIR ROSENZWEIG: Not enough yet?

MS. BAL: Still missing two more, if we could just try to click one more time.

Okay. We have high, 10; moderate, six; low, one; and insufficient evidence for three.

CO-CHAIR ROSENZWEIG: Okay. So the next issue is related to the performance gap that is addressed by this measure. And here, I think there is no question that over a period of time, the performance gap has narrowed. Would the reviewers like to comment

on this?

MEMBER McCOLLISTER-SLIPP: Yes. I mean, obviously, the performance gap has narrowed. There is, you know, in some health plans, like commercial health plans, there is significant compliance in meeting this quality.

But having said that, I would say that there is enough variation, especially if you look at Medicaid, that this still needs to be -- and even Medicare, the HMO rate for Medicare, I mean, it kind of blows my mind that somebody wouldn't be testing somebody who has diabetes for A1c. I mean, as, you know, just sort of a -- I'm not a huge fan of it as a measure, but it's -- if somebody is not doing that, there is a pretty good chance that it's not a particularly high-quality physician.

CO-CHAIR ROSENZWEIG: You know, I'm not actually as concerned about the Medicare rates. They are pretty high, and

there are a lot of very elderly people in that population that probably don't necessarily need yearly A1c rates. But the Medicaid certainly is, you know, 20 percent of the patient population is not getting A1c's on a regular basis. So I think that probably is more significant than I thought it would be.

Someone else had a comment over here? Bill.

CO-CHAIR GOLDEN: Yes. You know, in our state, it's worse. And some of it is -- and there may not be the position or the clinic. It could be the access issues and the outreach needs, especially with ACA and expansion of potential new patients under the systems with coverage. Unfortunately, it still has some validity and use. So it's -- we're using it now in dashboards for big systems, and it's -- there is still quite a bit of a performance gap.

CO-CHAIR ROSENZWEIG: Sue?

MEMBER KIRKMAN: So is the

performance gap narrowing or improving? I mean, it looks to me like since it was last endorsed the numbers have stayed about the same. I mean, Medicaid is lower. Everybody else is at about 90 percent over the three years. So is it really driving improvement in care at this point? Or is it kind of where it's going to be and it's not driving improvements?

MR. REHM: In that section that has our performance data, which is I think 1(b)(2), if you'll look at the 10th percentile unit, which is really the lowest bracket, you'll see that that is moving up, and it's moving up, it looks like about a point or a point and a half.

Generally, over all of our measures a point a year is not -- that's pretty good. It's actually moving a lot of populations into improved care. I think the other thing that we really don't understand here, and Robert, who is head of our Research

Unit, may throw a brick at me for asking to think about other things, but we will probably increase -- and I don't know the epidemiology, but my sense is that we are adding more people with diabetes into the denominator than we are, if you will, treating correctly.

And so you are chasing something that's ballooning. And to maintain a rate like that -- so I'm speculating because I don't have the data and I'm sorry. That could be a significant accomplishment. So sometimes there is more underneath the radar here than we may see from just the performance rates.

MEMBER KIRKMAN: But it seems like the 10th percentile in some groups it has gone down, in other groups it has gone up. Am I reading this data wrong?

MR. REHM: I was focused on the Medicaid, because that had been brought up as an important area. I think that has gone up.

MEMBER KIRKMAN: Okay. But like commercial HMO, it's actually gone down.

Commercial PPO, it's gone up. There is one where it went from 63 to 34 to 62 over three years.

DR. BARTON: I think the 34 is an error in that table.

MEMBER KIRKMAN: Okay. But, again, if the -- it went from 63 to 62. So I guess I don't really see this trend towards even the 10th percentile going up.

MR. SAUNDERS: I think one thing we might sort of emphasize is that -- so HEDIS, as it's implemented across these health plans -- Medicaid, Medicare, commercial -- is a pretty mature program, and many of the plans that are participating in this have been doing this for a while.

But the measure has uses outside of HEDIS, and that -- say, use in the exchanges or use in other contacts where you sought health plans that may be measuring populations for the first time, you may -- while it may be a low bar in some

circumstances, it will not be a low bar for those populations.

And so we might see greater performance gaps in other measured populations. It's just that we don't see that in our data because many of these plans have been working on this for so long.

MEMBER KIRKMAN: But I guess, again, and maybe I'm getting too meta here, but, you know, if you're spending time collecting this performance measure, you're not spending time on something else. So, you know, I mean, I just -- I don't see a clear pattern that it's improving care. Rates are pretty high, and it's embedded in another -- two other measures.

So, I mean, I think it is -- you know, you can't just say oh, it's okay, we can keep collecting it, because again, if you're collecting this, you're not doing something else. So, and there can be too many measures.

MEMBER McCOLLISTER-SLIPP: Is that

true, though? I mean, and I honestly don't know how this is done. I would think that that would be a relatively easy thing to extract from EHR data or other things that you are going to already be collecting.

MEMBER KIRKMAN: Well, even so, I mean, if -- it does take some time, even if it's pretty easy. And so if it's not worth doing, or if you're -- if you're not able to do something else that might be more worthwhile because you're doing this, then that seems like a reason not to do it. But --

CO-CHAIR ROSENZWEIG: There are two issues. There is whether or not there is enough of a gap, and then there is the issue of whether or not this gap is amenable to being improved. So the question I have -- I would have related to this is, what percentage of the patients who don't get A1c's measured are not seen in the previous year? Do you have any data on that?

DR. BARTON: We don't have data on

that. But I would say that, you know, the question is, is there -- is there a gap in care? Not is there a gap in care everywhere? And what -- you know, the data that we have provided to you shows that, you know, just, for example, the Medicaid HMO rate is very stable I the median, 77, 78, 79, but the 10th percentile has gone up from 41 to 59 percent.

So these are places that are, you know, organizing their care differently to get more -- I mean, at least you have to get the patient in if they are going to get their A1c tested. But we don't have data to answer your exact question.

CO-CHAIR GOLDEN: Okay. Can I make a comment about burden? I would say that this -- the burden of this measure is trivial for a couple of reasons. One, it's an administrative measure; people are familiar with it. Two, they have already developed the algorithms. So it is already -- it is just off-the-shelf software for most programs now,

so they just have to rerun the algorithm.

So I don't think it would be -- the amount of work involved now to replicate the measure year in and year out is pretty small.

MEMBER KIRKMAN: Well, again, I mean, I -- so I'm new to this process, but you do retire measures, right? You do sometimes drop them. I mean, even though you say -- I mean, I agree, it's easy, but it's kind of like we tell our primary care doctors, well, it's easy to just do one more thing or to do -- you know, follow one more guideline. Or, you know, this is not going to take that long, but again, it's the totality and I just think we should think carefully about whether this is providing enough benefit and enough additional information to continue it.

CO-CHAIR ROSENZWEIG: Any other -- yes, Ingrid.

MEMBER DUVA: I have a question for Bill Curry. Can you explain again -- you

said that the eight percent and the nine percent, the good care and the bad care, they are not going to provide the provider the information they need. I didn't understand that, because I thought you would need the measure just to calculate those measures. They are not available?

MEMBER CURRY: So if we look at the list of patients who have an A1c over nine percent, embedded in that population is a group of people that did not have an A1c in the past year. So it's going to be more difficult perhaps to be able to give that list to the provider to say, you know, here's a list of patients that have an A1c over nine. Or are they on that list because they weren't tested in that year?

So the process measure is embedded in both of those. But to provide a list of patients who did not get checked in the last year, as this one is doing, as an easy way for them to look at the gap and then have their

care team engage that patient in care, it makes it easier for those people that are using this kind of -- either claims data from our carriers or that we generate internally.

MEMBER DUVA: Okay. I just -- thanks. That clarified it. I just wanted to know that works on the assumption that you are not trying to, you know, take apart your process and figure out what is wrong with your process because you've got all these people with -- in the over nine percent category that may not be over.

Now, I know at the VA, we make a directed effort to get everybody tested, so we get them out of that. If they are erroneously in that, you know, denominator, then they are -- or in the numerator, then they're out. So I see what you're saying. Thank you.

CO-CHAIR ROSENZWEIG: Jessie.

MEMBER SULLIVAN: Yes. I guess I just wanted to offer Ingrid another explanation. So we're a health plan, and many

of -- about 60 percent of the providers in our network are small doctors in private practices in rural areas, and they don't -- many of them don't have EHRs and they are not doing measurements themselves.

So we give them lists of their failing members, and that allows them to do this quality. But they -- as Mary pointed out, the sample for the value of A1c is for 11 across the membership of our entire plan. So it might be one of the diabetics in your practice, but the denominator for this measure is everyone. So we are giving them a list of everyone who hasn't been in, whereas on that measure we'd just be giving them the one person who fell in the sample.

MEMBER DUVA: Okay. So how does that not cover just the standard A1c testing, then? I mean, you're still going to be attacking the same problem when you try to improve. You've got to get the test done, right? No?

MEMBER SULLIVAN: So I guess I'm saying that because this is a measure, we give doctors in our practice a list of their hundred patients with diabetes and which ones haven't been tested. If we were only doing the A1c level test, we would give them the name of the one person who fell in the measure in the same and had failed. So we are giving them a much smaller sample because this measures everybody.

DR. PACE: So did the other measures. The denominator was everybody, right?

MEMBER SULLIVAN: Not as it's implemented by NCQA.

MS. BAL: So we'll be voting on performance gap, which is -- hold on one second -- which will be data demonstrating considerable variation, overall less than optimal performance across providers and/or population groups. And you can vote now.

Okay. We have high, three;

moderate, 13; low, four.

CO-CHAIR ROSENZWEIG: Okay. Let's move on now to discuss the impact. Any comments?

MEMBER KIRKMAN: Can I ask a more general question about this priority category? It's not necessarily that the disease is a high priority. It's that the measure itself is a high priority, right? So, I mean, for -- if it's just the disease, we could say all of the diabetes ones are a high priority, right? So it's really the measure. Is that correct?

CO-CHAIR ROSENZWEIG: It's my understanding, yes, that it's the measure itself, not the -- obviously, not the disease. So it's the priority of this particular -- the impact of this measure on overall health. I mean, obviously, diabetes has a high priority. But whether or not this particular measure is going to influence or be associated with improved care -- yes.

MEMBER TAYLOR: And to put a

little finer point on that one, it's -- if I understand it right, it is in the context of having the other measures that are already out there right -- our less than nine and our less than eight. Incrementally, how much does this add, as a priority? Is that the correct way to see this question?

CO-CHAIR ROSENZWEIG: I believe so. Is that the general view?

DR. PACE: You know, we ask you to look at the measures independently, but, you know, certainly it is looking at the condition as well as the impact of poor quality on this particular measure, so -- or, on what is in the numerator. So it's a combination of those things, but I think you're right in -- you know, so I think in terms of your evaluation, we look at each measure independently.

So, you know, it really is to think about the target population as well as the numerator event or process that's being measured and what impact that has.

CO-CHAIR ROSENZWEIG: Okay. Any other comments?

(No audible response.)

Okay. Why don't we vote on priority, then.

MS. BAL: Okay. So high priority addresses a specific national health goal or priority, or data demonstrated a high-impact aspect of health care. And you can begin voting.

Okay. So we have high, eight; moderate, seven; low, five.

CO-CHAIR ROSENZWEIG: Quite a spread. Okay. So the next issue is the reliability of the specifications. Comments by the reviewers?

MEMBER McCOLLISTER-SLIPP: In terms of the specification or specificity, I mean, the workgroup certainly concluded that it was highly specific. Are we looking at that specifically or -- no pun intended. Reliability, I mean, there seems to be a

pretty high suggestion of reliability in terms of the ability to collect the data. And, I mean, I don't think there was much discussion around that. Do you remember, Bill? It has been a few weeks, but based on the comments here it seems to be relatively -- relative degree of certainty about the reliability.

CO-CHAIR ROSENZWEIG: Okay. Any comments by anyone else?

(No audible response.)

Let's vote on this one.

MS. BAL: Okay. So reliability would be the specifications and testing for this, and the voting is now open.

Can we try again? We're missing one person. This one has high, 16; moderate, four.

CO-CHAIR ROSENZWEIG: Okay. So let's go on to validity here.

MEMBER McCOLLISTER-SLIPP: Again, trying to remember back to where we were, I mean, some of the validity questions that were

raised by the workgroup were -- you know, would it -- does it make sense to include people under 18 or over 75? And I would say that there is -- I'm not completely sure why that determination was made.

And, I mean, the test, in and of itself, seems to be relatively valid. I think this gets back to some of the necessary but sufficient -- that the way the test is conducted seems to be valid.

Bill, any further comments? I just don't know how much detail you want me to get into.

CO-CHAIR ROSENZWEIG: No, I think that's okay.

MEMBER McCOLLISTER-SLIPP: Okay.

CO-CHAIR ROSENZWEIG: I think the issue about over 75 is -- I mean, obviously, a lot of people over 75 would benefit from A1c testing, but at a certain point you do get into a situation where, especially in patients who have relatively mild diabetes, whether or

not getting A1c's would necessarily be of that much benefit.

With respect to less than 18, I think we are just restricting ourselves to the adult population as far as this measure is concerned. And probably, you know, the pediatric population, you know, will have to have additional measures for them separately.

MEMBER HAYDON-GREATTING: I don't know that there's a reliability of the A1c past that age range because of the variability in the hemoglobin when you get into those complicated elderly patients. I mean, as an example, my mother-in-law runs 200s, 300s on her finger checks and her A1c came back last week at 6.5. So it should be showing higher for her if she's running these daily glucoses at 300s and 200s and -- but she's 86 years old. She gets a little -- she says she gets goofy, she sits down, she -- you know, I mean, so she's managing it, and I see a lot of elderly patients, especially little, mini,

frail people who the hemoglobin A1c may not be the -- and there's some study out there. There's somebody that just recently hit Newsweek about challenging -- you know, should we put all of our credit into the HbA1c as we are.

CO-CHAIR ROSENZWEIG: Well, this issue has come up, obviously -- oh, I'm sorry.

MEMBER KIRKMAN: It actually, on average, it tends to run a little higher in older people. But I don't -- but I think, you know, the other issue might be that, you know, at a certain point also, it might not be that they have such mild disease, but maybe somebody's got so much comorbidity. You know, an Alzheimer's patient in a nursing home, do they need an A1c? Probably not.

CO-CHAIR ROSENZWEIG: Yes, that was --

MEMBER KIRKMAN: That's not saying everybody over 75 falls into that, but --

CO-CHAIR ROSENZWEIG: Yes. As I

get older, I -- I mean, there are a lot of people -- yes, at a certain age, even if the A1c is accurate, is it really necessary? Of course, as I get closer to age 75, I think, oh, it must be much more important.

(Laughter.)

But the other issue that has come up that's really a separate issue from this, but it more relates to the issue of using A1c to diagnose diabetes, is that there has been a lot of controversy about that because of their different relationships between A1c and average blood glucose control in different ethnic populations, which has come up. East Asians, Indian populations, as well as Hispanic populations, as far as I can recall.

But I don't think that that necessarily applies here where you are looking for overall glycemic control in patients with diabetes.

Any other comments?

(No audible response.)

Okay. So let's vote on this.

MS. BAL: Okay. We're voting on if the specifications are consistent with evidence, and that testing and threats are addressed. There is -- the exclusions, risk adjustment, meaningful differences, multiple specifications, and missing data are all looked at. And then you can go ahead and vote now.

So we're just looking for two more. If we could just try and -- I guess try one more time. There we go. Thank you, guys. And so we have high, 11; moderate, nine.

CO-CHAIR ROSENZWEIG: Okay. Feasibility.

MEMBER McCOLLISTER-SLIPP: Oh, I found my place this time, so slightly less flipping around.

In terms of feasibility, there was general agreement this is a pretty feasible, pretty easy to access statistic, or pretty easy measure to extract from existing claims

data or EHR data, things that people are already collecting. So it didn't seem -- the workgroup did not seem to think it would be additional burden.

If there are processes -- I mean, I do EHR data stuff, so my sense is that it would be relatively easy to extract. But if there are processes that would affect smaller practices or other people, I'd love to know what those are. I don't know, but it seems like of the measures this would be pretty easy to come by.

CO-CHAIR ROSENZWEIG: Any comments?

(No response.)

Let's vote.

MS. BAL: So we're voting on feasibility, and that is for -- that the data generated during care, there is electronic sources, and data collection can be implemented. And the voting has started.

And the results are 18, high;

moderate, two.

CO-CHAIR ROSENZWEIG: Okay. So use and usability. This may have slightly more discussion. Yes.

MEMBER SHWIDE-SLAVIN: Well, I think there would be a possible unintended consequence if this was not done, because it has taken a long time to educate the public that they need to have an A1c test done to understand how their diabetes is doing and for the physicians to take the test. And if they weren't being measured, I wonder if that message would continue to be heard out there.

CO-CHAIR ROSENZWEIG: So, a good point. I mean, that has been applied in other situations. Yes. A lot of times the presence or absence of measures is used as a basis for denying care or essentially letting a plan say that we won't pay for this or that measure, this or that test. Obviously, A1c is not a very expensive test, but we will be getting into issues related to bone densitometry

later.

Any other comments?

MEMBER McCOLLISTER-SLIPP: No. I mean, that workgroup seemed to be -- I mean, putting the philosophical issues aside about whether or not A1c is a good test of quality or whatever, the existence of using this as a process measure was pretty unanimously accepted. I mean, it seems useful.

Again, if somebody is not doing an A1c test on one of their diabetic patients, I would question what -- you know, whether or not they were a competent physician. So, and I think the workgroup was in agreement with that, unless anybody remembers something I'm forgetting.

CO-CHAIR ROSENZWEIG: Okay. Let's vote on use and usability.

MS. BAL: All right. So for use and usability, we are looking at accountability, transparency, demonstrated improvement, and the benefits outweigh

evidence of unintended negativity, negative consequences. And it's open now.

And the results are high, 14; moderate, four; low, two.

CO-CHAIR ROSENZWEIG: Okay. So now we are going to vote on the overall recommendation for endorsement. Any comments first? I don't see why there should be, but --

(No response.)

Okay.

MS. BAL: Okay. Voting is now open for overall suitability.

Final results are yes, 18; no, two.

CO-CHAIR ROSENZWEIG: Okay. Thanks. Let's move on to the next one. Which one was that? It's going to be 0055, comprehensive diabetes care, eye exam, retinal eye exam performed. Who were the reviewers? Oh, sure. Absolutely.

DR. BARTON: Okay. Thanks very

much. The comprehensive diabetes set that NCQA uses to evaluate health plans includes this measure, 0055, which looks to see if those people who have diabetes, which is defined exactly the same as for the other indicators that you've seen, with the implication being, and in fact the practice being, that they collect all this information on one defined group of people.

The high risk for vision-threatening microvascular complications of diabetes is very well-known, and the opportunity for early intervention by an ophthalmologist to treat the kinds of microvascular events and hemorrhages in order to preserve vision is really the focus of this measure.

And the numerator of the measure can be complied with by seeing any eye care professional, so that includes optometrists as well as ophthalmologists and -- within the measurement year, or having had a normal or

negative exam the year before. And that's 0055.

CO-CHAIR ROSENZWEIG: In their definition, do they include retinal photos that are done remotely that might be read by a qualified eye person? Because that was -- that has always been an issue.

DR. BARTON: I'm sure if the reader bills the visit, it would meet the criteria for the code. Because this is something that uses claims to determine whether they had a visit with an optometrist or an ophthalmologist. So I don't know the particulars of that kind of distance care arrangement, but something tells me that the person reading would charge for that. And so it would get counted.

MEMBER McCOLLISTER-SLIPP: It was a formal interpretation that they would bill for.

CO-CHAIR ROSENZWEIG: Correct. I mean, a lot of -- there are a lot of

photographic telemedicine systems now that -- in which you can get the picture taken in the primary care office, and then it's sent to be read officially. So that would count, then, as part of this.

MS. TIGHE: Anna, do you want to start with evidence for this measure?

MEMBER McCOLLISTER-SLIPP: Sure. Sorry. I'm just having a hard time following where -- the printed worksheet. So I think this is maybe one big massive blond moment. I'm sorry.

Okay. Here we go. Thank you very much. So the evidence for this seems to be pretty strong. I mean, one question I had about -- in sort of the philosophical discussion we had during the workgroup, the call was who was actually being measured for this. Is it health plans? Is it physicians? Is it my endocrinologist? Is it the ophthalmologist? Because I think that will matter substantially.

And I know the numerator/denominator statement, you know, talks about the patient specifically, but, you know, I don't necessarily think it's appropriate to hold my endocrinologist accountable for whether or not I make it to the ophthalmologist or not for a dilated eye exam. And given the fact that there are much improved point of care retinal exams that could be given in the primary care setting, I don't think that's particularly ubiquitous. So expecting primary care physicians or endocrinologists to be able to do that would probably be inappropriate, at this point in time at least.

So I guess the question -- we discussed this a bit in the workgroup -- that I would have in terms of the measure is, what was the rationale for that? And is there some degree of specificity on this that at least I haven't seen, in terms of who is being measured.

CO-CHAIR ROSENZWEIG: Well, obviously, you couldn't hold the ophthalmologist or the person reading these responsible for the percentage of patients who are actually read, because obviously the denominator includes all people with diabetes within that certain age group.

So it would either have -- if you're talking about it, it would either have to -- the responsibility would either have to be on the plan level or on the primary care level to a certain extent. And I guess primary care doctors are being held responsible for sending their patients to the ophthalmologist. If they don't get to the ophthalmologist, that is a valid issue. But the issue is also that -- whether or not -- how vigorous the individual person is or the system is in getting the person to be tested. Yes?

MEMBER BREEN: Just it's an interesting discussion, because when we get

into medication adherence, right, and how practices are being measured on that, at first blush clinicians may say, "Well, you can't measure me on medication adherence. It's not my problem if my patient doesn't take their meds."

However, when you really begin to look at that data, you see major practice variability amongst the similar demographic of people that says there are ways that you can structure your practice, right, to deliver better care. And I think closing this loop on the eye exam is one of the things that primary care should be challenged to do, and endocrinologists as well, because there is a range of activity that I can do when I have a patient. I never tell you to go see the eye doctor. I tell you to go see the doctor. I tell you to go see the eye doctor and I write you a referral slip. I tell you to go see the eye doctor, I write a referral slip, and my secretary calls the ophthalmologist and books

you while you're right there versus -- you know, I mean, you see where the spectrum is.

So I do think it's a valid measure from a clinical standpoint. The burden is on us to just do this.

CO-CHAIR ROSENZWEIG: Okay. Let's try to focus on evidence specifically. Yes.

MEMBER CURRY: I just wanted to comment about that there is no specificity in the measure that says that a mydriatic or non-mydriatic digital photo of the retina meets the definition in here. You know, our region -- our insurers will not cover that. They have to have a visit in the optometrist's or ophthalmologist's office. We have tried to do this in our rural practices and in our academic practices, and they will not accept that.

So it -- the specific language of digital retinal photograph is not in there, so they will not cover it.

CO-CHAIR ROSENZWEIG: That doesn't

mean that they're right.

MEMBER CURRY: I'm just saying --

CO-CHAIR ROSENZWEIG: Okay. Okay. Someone else? Oh, yes. Oh, but one other issue that does pertain to evidence is that there are a number of studies that show in-patients with diabetes who are very well controlled, that they don't necessarily need yearly retinal exams, that they could go every two years. There are papers by Joe Selby, I think Carol Mangione, and several other papers that -- but the issue -- yes?

MEMBER KIRKMAN: There's evidence that people can go for three years if they had a negative exam.

CO-CHAIR ROSENZWEIG: Correct.

MEMBER KIRKMAN: Including in the Medicare age population.

CO-CHAIR GOLDEN: So to follow up on that, that is a concern I was going to raise is, you know, obviously, a yearly standard drives costs and inconvenience and

measurement. And if the evidence now shows you can do it less often, then is this measure appropriate? And so the question is, what is the evidence for frequency? Do we have that?

CO-CHAIR ROSENZWEIG: Well, the evidence certainly is there for patients who are not well controlled, but the -- but for patients who are extremely well controlled --

MEMBER KIRKMAN: Well, I think it has to do with your -- the findings on your initial retinal exam. So there is -- I mean, there is even fundus photography evidence that if you had a normal fundus photograph you can go three -- people can go three years before their next one with no difference in outcome.

So I'm concerned that this is a little bit more aggressive than the evidence would suggest.

MEMBER MILLER: Concerning the photographic exams, the ADA guidelines say that the photo is okay periodically, but the American Association of Ophthalmology says

that it's of limited value for very early detection and diagnosis.

CO-CHAIR ROSENZWEIG: Yes. For Type 1 diabetes, it is usually not recommended for the first four years. It used to be five years, but it's more like --

MEMBER MILLER: Yes. Three to five for people who are initially diagnosed. But, I mean, for initial diagnosis of retinopathy they're saying that it's not always the best, the camera.

CO-CHAIR ROSENZWEIG: Oh. You're talking about the camera specifically.

MEMBER MILLER: Yes.

CO-CHAIR ROSENZWEIG: Oh.

CO-CHAIR GOLDEN: So to follow up on Sue's comment before the NCQA comments, you know, the question is, is this a measure for screening? Which would indicate a certain frequency. And if you already have an existent disease, it is no longer screening.

So that would require a more

intensive followup. And are we mixing apples and oranges in how we construct a measure and the frequency?

MR. REHM: Just to read from the denominator, the patient is numerator-compliant if the eye exam was performed or a negative eye exam was documented in the year prior to the measurement year. So it's more than just the measurement year. If you have a negative finding, then -- so --

MEMBER KIRKMAN: But I can tell you that every letter I get from Aetna or, you know, other people it's basically your patient hasn't had an exam this year. So I'm not sure it's really being -- that is really being adhered to. But even the two years is probably stricter than the data.

CO-CHAIR ROSENZWEIG: Also, the measure specifies screening in the numerator. But it doesn't specify patients who have existing retinopathy as an exclusion in the

denominator. That's a point that should be looked into, since by -- you know, after 10 or 15 years, the majority of patients with diabetes have some retinopathy.

So you're talking about actually a fairly significant number of patients in the population that may not necessarily need to be dealt with in this particular measure.

DR. BARTON: The workgroup brought up that point, and I think it's an excellent one. That this -- and as we further look at these measures, we will be looking at those things, both the interval and the question of existing disease.

I guess the issue about existing disease -- and it reminds me that this is a measure that is, you know, best used in a population of people. And when you are comparing one entity, like one health plan to another health plan, the likelihood is that they each have similar proportions of patients in those various -- either the people who only

need every three years versus every two, and the people who need more often because they've got eye disease. And so the -- drawing a line somewhere, you know, is the way the measure works.

CO-CHAIR ROSENZWEIG: But let's just talk about -- we want to focus on the -- I mean, this measure may be a useful measure for people to do to encourage eye screening. But right now we are talking about the evidence base for it. So --

CO-CHAIR GOLDEN: So let me just make sure I understand what you just said. You seem to indicate you are going to revise this or it should be revised. Should we pull this now and have you come back in six months?

DR. BARTON: The cycle on which we work is not that rapid.

(Laughter.)

So --

CO-CHAIR GOLDEN: What if we encourage you?

DR. BARTON: So we -- as you can imagine, our diabetes set, which also includes indicators that you are not seeing today that have to do with LDL screening and control, are going to keep us pretty busy over the next six months?

So we are -- I was indicating points that we wanted to keep in the queue for when we reevaluate this measure the next time.

CO-CHAIR GOLDEN: So you're saying we should endorse a measure that you think needs to be revised.

MR. REHM: Having been through this in other groups, other measurement domains, multiple times, we annually update our measures, and we get feedback from the marketplace thousands and thousands of either happy physicians or happy health plans or unhappy health plans and unhappy physicians, depending on the measure.

And we are constantly revising these measures, and we -- NQF has a terrific

process of -- because we have over 100 measures in play of doing quarterly updates. We update these with new code sets. We update these based on new technology. We update them.

So to think that a measure is static from the moment you endorse it is not fair. It is a very dynamic -- whether NQF was here or not, it's very dynamic in our world. It's so dynamic we put out technical updates in October, even though we reduced the measure specs in July, to capture the latest and greatest.

We don't release the NDC codes for any measures with drugs until late November, to capture the very, very last update of that -- you know, that compendium. So many times we bring measures where, because of timing issues, we are actually in the middle of evaluating the measure.

When we brought breast cancer in three years ago, we were right in the middle

of the evaluation. We couldn't say where we were going to land. And, you know -- and so sometimes it is just a timing issue.

I think the openness to change on any measure is a great thing, and, you know, we are not trying to dig out feet in the sand. It's just for something like intervals, we get feedback from you which is quite helpful. It's one of the benefits of participating in the process is that -- I hate to say this, but this is a free measurement advisory panel to help us develop our measures as well.

So you add that up with our own panels, and we bring back the feedback, that's part of the cycle of measure development and refinement. So I wouldn't say that the best solution to this, because we may be unsure about intervals, is to say that not endorsing the measure is the best thing to go -- but, you know, that's your decision obviously.

CO-CHAIR ROSENZWEIG: Just to get back, we are not discussing as to whether or

not this measure should be endorsed. We are -- let's focus, at least for the present, on evidence related to the measure. Anyone? Janice, do you have a comment? Sue?

MEMBER KIRKMAN: Yes. And, again, maybe I'm getting ahead of the evidence a little bit, too. But, I mean, I do think this is an example where we probably shouldn't let the perfect get in the way of the good. I mean, I think this is a measure that has been around a long time. I think it has done a lot of good. You can see there are still a lot of gaps in care. You know, I don't think it's all just because there are a lot of people with normal exams that aren't being referred back every year or two.

So, you know, I think, you know, in the future it would be good to tweak it a little bit maybe, but -- and the other thing is I think it becomes really difficult when you get different intervals for different people, and it is hard to tell from a -- you

know, an administrative level or SARP review level who is supposed to be at what interval. So, I mean, it --

CO-CHAIR GOLDEN: But wearing my Medicaid Medical Director hat, and speaking for other Medicaid Medical Directors, that is a cost item. And so it actually -- this has a big impact. So it does make a difference.

MEMBER McCOLLISTER-SLIPP: What's the potential cost of missing a vitreous hemorrhage? I mean, you know, every other year it really isn't that extreme, if that's -- if you're looking at cost issues. I mean, and I went in for a retinal exam and happened to be having a hemorrhage, and because I was going in for my regular exam while this happened I still have 20/20 vision.

So I think it's incredibly important, and I purposely go to a physician who is in like Southeast D.C., so I can see what it's like for other people for whom this is a huge burden to get there and to get back.

And it's sometimes a family project.

But, I mean, if we're looking at this from a cost perspective, the relative cost of requiring, you know -- using it as a quality measure coming in at least once every other year versus the cost of potential blindness or other complications -- or other types of surgery as a result of not doing it, I think the cost-benefit would probably weigh in favor of doing it.

CO-CHAIR ROSENZWEIG: I would mention, yes, I work in a safety net hospital with lots of people with disparities. And basically we have terrible eye screening rates. I mean, we have done all sorts of things to try to encourage the patients to get screened. We send them to the ophthalmologist; they don't show. A variety of other things.

But we also -- also identify a lot of patients with very severe eye disease at the very first -- at the very first interval,

so -- where we have missed the boat. So I think this is an issue that is -- it has certainly been raised by my colleagues at the Beetham Institute at Joslin that there are -- you know, that if you cut back on emphasis on screening that you could end up with a lot worse disease.

CO-CHAIR GOLDEN: Let me just clarify my comment, just so I could make it -- that doesn't bring it up on cost. But as a Medical Director looking at expenses for a program, if the evidence doesn't justify the frequency, then it's not a necessary expense. There are other things to spend the money on.

So I'm not saying just cut it back because of cost, but the question is, what does the evidence justify? That was my question and my point. So if there's questions about the evidence justifying mass screening on a yearly basis, then it has cost implications.

CO-CHAIR ROSENZWEIG: Yes. Bill?

MEMBER TAYLOR: Procedurally, what is the option for us if we believe in retinopathy screening and we think that this can be tweaked and improved? What is the most expeditious way for us to help get there if that's what we, as a group, conclude?

CO-CHAIR ROSENZWEIG: That's a good question.

MS. TIGHE: So we are asking you to vote on the measure as it is presented today. The tweaks that you all are suggesting are major changes to the specifications of the measure. It is not something that we could do through our annual update process. So the developers certainly can take that feedback as they are revising the measure, but you do need to look at it as specified today.

CO-CHAIR ROSENZWEIG: Okay. Let's have a vote on the evidence.

MS. BAL: Okay. So we are voting on evidence which should be -- the following should be considered -- quality, quantity, and

consistency, graded guidelines, empirical evidence, and expert opinion. Voting is now open.

Let's just all try one more time, get that last person in. Okay. The final results are high, four; moderate, 12; low, four.

CO-CHAIR ROSENZWEIG: Okay. Performance gap. Anna?

MEMBER McCOLLISTER-SLIPP: Based on the data presented, there seems to be a somewhat frightening level of performance gap from my perspective. So there certainly seems to be a need to emphasize this as something that should be done, you know, given the potential morbidity associated with not doing it I think.

And, again, other workgroup members, if there is anything I'm missing from our discussion, please let me know.

MEMBER MILLER: There was also somewhere in the developer's evidence,

discussion about disparities, that lower income patients were also less likely to receive eye exams. Just putting that out there, as far as the performance gap.

CO-CHAIR ROSENZWEIG: Yes. There's a lot of evidence. Okay. So let's vote.

MS. BAL: Okay. So we're voting on performance gap, which is data demonstrate a considerable variation or overall less than optimal performance across providers and/or population groups. And voting is now open.

Okay. So we're at high, 18; moderate, two.

CO-CHAIR ROSENZWEIG: Okay. Impact. Anna?

MEMBER McCOLLISTER-SLIPP: Again, I think -- I am completely incapable of finding myself through this worksheet, so my apologies for that. I don't quite understand why it's so baffling. But the potential impact for screening in terms of morbidity was

pretty high. And, again, we didn't have a long and extensive discussion about this on the workgroup call. But there seemed to be pretty significant agreement that the impact of doing this was potentially beneficial. Anyone?

MEMBER BREEN: Yes. I'd just like to state for the record that diabetes remains the leading cause of blindness in the United States. So just to get that out there.

CO-CHAIR ROSENZWEIG: Yes. I think in this case we are also considering the priority of the use of the measure, and we are not necessarily considering whether or not one or two years or three years is the best period of time. It's whether or not the measure itself might have a high priority or impact.

So, any other comments?

(No response.)

Okay.

MS. BAL: Okay. Voting is now open for high priority.

Let's just all vote again. There's a little delay I guess. We are at 18.

(Simultaneous speaking.)

Okay. High, 15; moderate, three; insignificant, three.

CO-CHAIR ROSENZWEIG: Okay. Reliability of the specifications.

MEMBER McCOLLISTER-SLIPP: In working from memory here, I think the only question we had around reliability was -- gets back to who was being measured, I mean, and how it was going to be recorded. So if you're looking at EHR data from ophthalmologists, that's pretty reliable.

If you're looking at primary care, endocrinologists, you rely on physician in many cases getting a letter from the ophthalmologist saying that the patient was seen or you're relying on the patient's memory or relying on the patient to be honest about whether or not they did it when they're embarrassed with their physician. So --

CO-CHAIR ROSENZWEIG: Any comments?

MEMBER McCOLLISTER-SLIPP: Bill, do you have anything else to add?

(No response.)

CO-CHAIR ROSENZWEIG: I think the major issue is just the -- with respect to specifications is the conflating of whether or not you are dealing with screening or you are dealing with following up of existing retinopathy.

Is the specification claims data or is it chart data? It's claims data? Okay. Thank you.

Okay. So let's vote.

MEMBER DUVA: Can I just add a comment?

CO-CHAIR ROSENZWEIG: Yes. Oh, I missed something. Sorry. Okay. Ingrid.

MEMBER DUVA: Sorry. The developer did submit reliability data, and they tested high.

CO-CHAIR ROSENZWEIG: Yes. Jessie.

MEMBER SULLIVAN: I'm sorry. I have a clarification. Maybe it's an implementation issue. But I believe this is a hybrid measure. We certainly spend a lot of time looking for those ophthalmology charts.

MR. REHM: It's a choice. There are plans that feel that their admin records are sufficient, and it's because of that range of systems -- some people have legacy systems -- they can do three-quarters of it. But then for the rest of it they take over new plans or -- so that's why we have that option.

CO-CHAIR ROSENZWEIG: Okay. Let's vote.

MS. BAL: Okay. Voting is open.

Okay. We have high, seven; moderate, 13.

CO-CHAIR ROSENZWEIG: Okay. Validity of the specifications.

MEMBER McCOLLISTER-SLIPP: The

validity of the specifications -- again, we didn't really cover this that much during our workgroup, because we were a little rushed. But, I mean, they seem to be valid just in terms of a process measure of doing the screening.

MEMBER MILLER: There seemed to be some high correlation for health plan level data, but the correlation was a little bit weaker for the physician level data, so with the diabetes recognition programs. But when I wrote some notes to myself, I said that it may be reflective more of the difference in sample sizes between the levels of data, between the health plan data and the diabetes recognition plan.

CO-CHAIR ROSENZWEIG: Yes.

MEMBER DUVA: Also, same for this measure as previous. The face validity was presented from experts, and the correlation to the other quality measures was high.

CO-CHAIR ROSENZWEIG: And is there

an agreement about whether or not a score from this measure is -- as specified is an indicator of quality? Screening rates in general?

MEMBER DUVA: Do you mean from the workgroup?

CO-CHAIR ROSENZWEIG: From the workgroup, yes.

MEMBER DUVA: In general, the conversation was that it was representative of the quality of care for the patient to be screened.

CO-CHAIR ROSENZWEIG: So let's vote on this, then.

MS. BAL: Okay. Voting is open.

Okay. The results are high, six; moderate, 13; low, one.

CO-CHAIR ROSENZWEIG: Okay. Feasibility.

MEMBER McCOLLISTER-SLIPP: I think feasibility was the area that we had the most discussion about, just because we don't live

in a system with comprehensive care. You know, people go to different physicians for different things, so, you know, again, measuring -- an endocrinologist or a primary care physician or a facility where that's given, using this as a process measure for the quality of care, while -- the fact that it has to be done by somebody else externally I think is an issue in terms of feasibility.

In terms of collecting the data, it is pretty straightforward, but -- from claims data. But in terms of feasibility, there seemed to be significant questions about how feasible it was to require them.

CO-CHAIR ROSENZWEIG: Certainly, data collection from -- you know, on a plan level seems quite feasible. But if one were to go into our -- like our medical charts to find out whether or not our individual patients had an eye exam within the past year, it becomes very difficult. So --

MEMBER MILLER: I think this is

one of the most difficult to measure variables outside of administrative plan data, because most EMRs don't have a physical field to collect to say that the -- or a lot of them don't -- to say that an eye exam was done and when it was done. And we can click that we referred the patient, but that doesn't mean that the patient actually went. So I think that on a provider or practice level I think this is extremely difficult data to capture.

MEMBER MAKAROFF: Yes. I would just agree with that. And also, I'm just wondering, it seems like from a feasibility standpoint -- and some patients may not use their medical insurance to get their eye exam, if they see an optometrist, use their vision insurance. Since it's not billed the same, then we're not going to capture that.

MEMBER MILLER: I have also over the years had many, many patients who go to Sam's Club and pay cash for their eye exam, and a letter never gets generated, you know,

and sent in. So --

CO-CHAIR GOLDEN: I have a question on the Sam's Club. I am not aware that they do diabetic retinopathy screening. They can do refractions, but --

MEMBER MILLER: I don't know. I have never personally availed myself of that.

CO-CHAIR GOLDEN: Yes. That is part of -- people think they are getting an eye exam and all they're getting is a refraction. So that's --

CO-CHAIR ROSENZWEIG: There are -- like Pearle Vision, you know, some of those places actually do do retinal screenings as a part of it. But it varies from place to place, and the patient doesn't know the difference for the most part.

So let's vote on this one.

MS. BAL: Okay. Voting is open.

So we have high, two; moderate, 13; low, five.

CO-CHAIR ROSENZWEIG: Yes. Sue?

MEMBER KIRKMAN: Do I remember correctly from this morning that another group is looking at a measure about the eye care professional communicating with the referring doctor? I mean, is there a sense that that may help some of this feasibility? Or is it so different that it's not?

CO-CHAIR ROSENZWEIG: I think that PCPI have a number of measures where they expect the communication from a professional to a primary care doctor. They are doing that with mammography, and I think they have done that with the eye exam.

DR. BURSTIN: There's a measure of patients who have retinopathy. Was there communication between the ophthalmologist and the primary care clinician? I think it was if they ever -- I mean, we can pull it up for you.

MEMBER MAKAROFF: I think there is also a measure in the CMS of being able to use core set for Stage 2 that is closing the

referral loop for all specialty care. So that would go along with that.

CO-CHAIR ROSENZWEIG: Oh, really.

Okay. Usability and use, accountability, transparency. Can we vote?

MS. BAL: Voting is open.

Okay. The results are high, seven; moderate, 11; low, two.

CO-CHAIR ROSENZWEIG: Okay. So let's vote on the overall measure. I hope you -- I'm not pushing you to move too fast. Okay. Sorry? Oh, I'm sorry.

MEMBER DUVA: Sorry. It's one of those questions, again, about kind of our group process. But somebody brought up that this measure might be kind of covered in other measures. But this is very close to a measure and coordination, and I know there is a coordination panel.

We look at care coordination separate -- with separate measures, but this is really a care coordination measure. And

I'm just wondering if we'll have an opportunity to discuss that later. It's about closing the loop. Really, the problems that we have talked about is, can we get the -- can we find out if the patient is -- can we get the data? So do we have an opportunity to talk about that later or not?

CO-CHAIR ROSENZWEIG: I don't know.

MEMBER DUVA: With these other measures?

CO-CHAIR ROSENZWEIG: I mean, I don't see this necessarily as a --

MEMBER DUVA: You don't?

CO-CHAIR ROSENZWEIG: -- coordination measure. It is basically among a specific population of patients whether or not they're in a plan or whether or not -- whether or not they are under the care of a specific provider, whether or not they get their eyes checked.

MEMBER DUVA: Right. But it's not

the provider providing the exam. So then they are going somewhere else to get the exam or the screening, and then it's coming back to their provider to know that they had it.

CO-CHAIR ROSENZWEIG: Yes. But we are not actually -- this measure is not actually measuring whether or not the provider is being told.

MEMBER DUVA: Okay.

CO-CHAIR ROSENZWEIG: You have to take that into consideration when you're, you know, voting on the overall value of the measure. I mean, it's -- you're making a valid point, but the measure is not specifically measuring closing the loop. At least as I see it here, the way it's written. It would be if you were just looking at the physician's chart, but that's not the case here.

Okay. Let's vote on this.

MS. BAL: Voting is open.

Okay. The final results are yes,

18; no, two.

CO-CHAIR ROSENZWEIG: Okay. Let's take a quick break. Five minutes.

(Whereupon, the above-entitled matter went off the record at 3:08 p.m. and resumed at 3:16 p.m.)

CO-CHAIR GOLDEN: Okay. So, let's continue with measure Number 0062.

PARTICIPANT: Excuse me. Would it be possible to have an update on the agenda and what you expect to cover today?

MS. TIGHE: Yeah, absolutely. This is Lindsey from NQF. I'll just update you.

We're going to cover Measure 0062 now, the NCQA Nephropathy Measure. From there, we're going to move to the three Joint Commission numbers 2416, 2417 and 2418.

After that, we'll move back to the foot exam measures. So, we'll be doing 0056, 0416, 0417 and 0519.

We are intending to cover every

measure that's on the agenda today. We are going to remove the 4:30 p.m. harmonization discussion to give ourselves a little bit more time back and hopefully still get out of here as close to 5:30 as possible. Maybe 6:00.

PARTICIPANT: Thank you.

CO-CHAIR GOLDEN: Okay. So, this measure is titled "Comprehensive Diabetes Care: Medical Attention for Nephropathy."

Bill, are you the major discussion well, wait. Do we want to hear from the developers first? Sure.

DR. BARTON: So, this is the final measure today from the Comprehensive Diabetes Care Measure Set.

Renal disease is another of the important downstream complications of diabetes, as I'm sure you all know, with really enormous impact on patients and families and the cost to the healthcare system.

Diabetes was the cause of nearly

half of new cases of end-stage renal disease in 2008.

There are many ways to enter this measure, that is to be numerator compliant, you could have appropriate laboratory screening for urinary protein, you could be referred to a nephrologist, or you could have evidence of treatment for diabetic nephropathy with an ACE inhibitor or an angiotensin receptor blocker. So, that's it's got a multi-prong way to be compliant.

Thanks.

CO-CHAIR GOLDEN: Has this measure changed in its specifications in the last few years? Because at least as I had heard the measure in the past, sometimes existing retinopathy has been used as a denominator exclusion.

DR. BARTON: Yes, I believe you're right. Let me double-check right now. We've got the specs.

Existing nephropathy

CO-CHAIR GOLDEN: Existing nephropathy

DR. BARTON: You said retinopathy, but

CO-CHAIR GOLDEN: has been used as a denominator and exclusion at least in previous versions of this type of measure.

This may be different, but in here you're including it really as a part of the numerator, existing nephropathy. So, I just wanted to check about that or is that a separate measure or is this a new measure?

DR. BARTON: Let me double-check that.

CO-CHAIR GOLDEN: Just a technical question, also, for a test for a would a metabolic profile with calculation of glomerular filtration rate count, or that does not count?

DR. BARTON: Does not.

CO-CHAIR GOLDEN: And why doesn't it count?

DR. BARTON: The test is for urinary protein burden. So, it is looking I think it's a 24-hour urine collection if I'm not

CO-CHAIR GOLDEN: Okay. But you're talking the label of this was "nephropathy" as opposed to urinary protein excretion. So, it's different.

DR. BARTON: So, you would argue that there's a glomerular filtration rate threshold that you would use?

CO-CHAIR GOLDEN: Well, you're not in Stage 1 renal failure. You can do that easily by a blood test with calculation of GFR.

DR. BARTON: Thanks.

CO-CHAIR ROSENZWEIG: The EGFR would be a separate measure which actually has been approved by the that joint NCQA-AMA panel recently.

I don't know if it's ready for submission to us, but the use of EGFR as a

regular tool has I know has been approved by your joint committee, but clearly it's not a part of this measure.

Yes.

MEMBER BAILEY: And just to add to that, that there is a significant proportion of diabetic patients with established chronic kidney disease that don't have overt proteinuria let alone microalbuminuria, or further along in the spectrum it's about 30 to 40 percent.

So, by omitting EGFR, we may be missing a significant portion of the population and also a significant opportunity to direct them to care and, hence, impact outcomes.

CO-CHAIR ROSENZWEIG: Yeah, that's why I was asking about the way this was constructed, because it was my understanding that the purpose of this particular measure was to promote the use of microalbumin as a screening tool.

It wouldn't necessarily diagnose nephropathy, but it would diagnose evidence potentially leading towards the diagnosis.

MEMBER BAILEY: Right, but you'd also be missing patients that may have decreased GFR and not have microalbuminuria. So, you should have either/or, or both.

CO-CHAIR ROSENZWEIG: You mean a composite measure, yeah. Right now as I understand it, they exist as two complementary measures, but I don't know if they've been put together as a composite measure.

DR. BARTON: PCPI measure is a physician-level measure and is not yet at the point of being tested.

CO-CHAIR ROSENZWEIG: Okay.

DR. BARTON: And so, it has not been brought to NQF yet.

CO-CHAIR ROSENZWEIG: Okay. So, were you able to clarify that issue about numerator versus denominator just before we had the discussion?

MEMBER KIRKMAN: So, I believe it's in the numerator. So

CO-CHAIR ROSENZWEIG: Here, it's in the numerator.

MEMBER KIRKMAN: Right.

CO-CHAIR ROSENZWEIG: But in

MEMBER KIRKMAN: But that's the program.

CO-CHAIR ROSENZWEIG: other versions of this measure I've seen it in the denominator. In other words, patients who already have existing nephropathy don't need to get microalbumin screening.

MEMBER KIRKMAN: But that would sort of that's what would happen here as well.

CO-CHAIR ROSENZWEIG: No, it's in the numerator.

MEMBER KIRKMAN: Right. They either get the test, or they have diagnosed nephropathy.

CO-CHAIR ROSENZWEIG: Okay.

MR. REHM: Just to clarify again on Section S4 Numerator Statement, patients who received nephropathy screening tests or had evidence of nephropathy during the measurement year.

CO-CHAIR ROSENZWEIG: Okay. So, it's basically trying to identify people who have evidence for nephropathy, but it's not but your point is well-taken that it's not clarifying whether EGFR is being measured, which would be the third piece to that.

Could you give your evaluation, please?

MEMBER TAYLOR: Yes. Our subcommittee spent most of its time on other measures, but generally went through this and was very favorable toward it. I guess we'll go through the specifics as we scroll down.

The general background is what Mary has already said that the evidence is good, that it is important to find. And that if you find it early, you can actually change

the course of events so that it makes sense to be criterion for screening with a big performance gap in disparities involved as well.

High priority for the reasons that Mary said with lots of chronic renal disease ascribable to diabetes.

CO-CHAIR ROSENZWEIG: So, the evidence quality?

MEMBER TAYLOR: Consider it high-quality evidence by the Subcommittee.

CO-CHAIR ROSENZWEIG: Comments?

MEMBER MILLER: I have a question and it goes back to the microalbumin EGFR. I always understood the urine microalbumin test to be a test to be a very, very early detection and identification of patients.

So, I guess the question, and I don't know the answer to this is, is microalbumin I know about the 30 to 40 percent who don't have microalbumin. But in those who do, does that generally, I mean, I

think that generally happens before their GFR would drop.

CO-CHAIR ROSENZWEIG: Yes. You actually have a rise in GFR for several can be for several years. You have an abnormally high GFR that often occurs in the very early stages of diabetic nephropathy.

So GFR, per se, a decrease in GFR really is detecting nephropathy at a further point than microalbuminuria might be.

But microalbuminuria but a lot of people get nephropathy without microalbuminuria, as Bob has just indicated. Maybe 30 percent.

MEMBER MILLER: I just was bringing that up, you know, because of the discussion we were having about if GFR is included or not that

MEMBER BAILEY: If I could just add to that, so microalbuminuria identifies the high-risk group that's more likely to progress down the spectrum than the general population.

MEMBER MILLER: Absolutely.

CO-CHAIR ROSENZWEIG: When the microalbumin test was first used, we kind of thought of it as being an indicator of getting nephropathy, but now it's really considered early nephropathy.

MEMBER MILLER: It is.

CO-CHAIR ROSENZWEIG: Yeah.

Any other comments?

MEMBER TAYLOR: We can keep scrolling.

CO-CHAIR ROSENZWEIG: Okay. So, let's vote on this then. This one is an easy one to vote on.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. The final vote is high, 13. Moderate, seven.

CO-CHAIR ROSENZWEIG: Okay. Performance gap. This is I think this is something that's actually improved over the years, but I don't know if it's

MEMBER TAYLOR: Yes. We didn't show the numbers right here, but there's a lot of gap in general in the population, and then disparities in particular high-risk groups.

MEMBER MILLER: I had made a note that the Medicare HMOs had the least improvement, but they had the highest mean percentages of performance.

CO-CHAIR GOLDEN: Let me ask a question on that. In terms of the gap, I find that just intrinsically hard to believe.

I guess my question is, is that because they are collecting the measure incorrectly?

My data would show that 80 percent plus of diabetics run ACEs and ARBs. So, your potential for a tremendous performance gap would be fairly low.

So, I was just curious if they're just measuring urine microalbuminuria, then sure. But if they're not excluding or accepting ACEs and ARBs as part of the

collection of the data, there would be a lot of variation.

MEMBER MILLER: I would think so. I think in a number of primary care practices there is a lot of emphasis on get the patient's urine before they leave, because that's something we can collect in the office even if the patient is going to a lab somewhere, but we can get a urine specimen.

But that really would raise all boats, not just the

CO-CHAIR GOLDEN: Yeah, but if your patient's already on an ACE or an ARB, you satisfy the measure.

MEMBER MILLER: That's right.

CO-CHAIR GOLDEN: That's why I'm confused by the performance gap.

MEMBER MILLER: I see. I don't have an answer for you.

CO-CHAIR ROSENZWEIG: All right. Let's vote on the performance gap.

MS. BAL: Voting is now open.

(Pause.)

MS. BAL: Okay. The final results are high, 11. Moderate, seven. Low, two.

CO-CHAIR ROSENZWEIG: I should mention that if a person is on an ACE or an ARB when they have diabetes, it's not necessarily an indicator and a fact that nephropathy is being treated.

It could be that they're being treated for hypertension. Okay. So, that's just something that needs to be considered.

CO-CHAIR GOLDEN: True. But if you're positive for urine microalbuminuria, it makes no difference.

CO-CHAIR ROSENZWEIG: No, but the numerator is saying microalbumin, or being on an ACE and an ARB, or being referred to a nephrologist. At least that's the way I read the measure.

So, there may be some people who are on an ACE and an ARB purely for hypertension who may not necessarily have

nephropathy.

I don't know whether it really matters in terms of the value of the measure itself, but I just thought I should mention it.

Yes.

MEMBER BAILEY: There are a couple of things. The first one is that very often the ACE or ARB is titrated upwards until you get no further decrease in urine protein or the patient can't tolerate it.

So, from a clinical perspective there are differences. But I think from an administrative dataset perspective, there probably was not.

CO-CHAIR ROSENZWEIG: Any other comments?

MEMBER MILLER: I don't see ACE and ARBs in the numerator here. I just see patients who receive nephropathy screening tests or had evidence of nephropathy, but I don't see

CO-CHAIR GOLDEN: Page 16 of the document.

MEMBER MILLER: Okay. Thank you.

PARTICIPANT: Also at the very end is an algorithm.

CO-CHAIR ROSENZWEIG: Yeah, they're defining evidence of nephropathy as including being on an ACE and an ARB, which is not exactly technically correct.

MEMBER MILLER: Excellent. Thank you.

CO-CHAIR ROSENZWEIG: Any other questions or any other comments?

(No response.)

CO-CHAIR ROSENZWEIG: Okay. Let's vote on priority then.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. It's high, 16. Moderate, four.

CO-CHAIR ROSENZWEIG: Okay. From the Workgroup, any comments about reliability?

MEMBER TAYLOR: Yeah, only that there was some evidence that showed that the test is reliable.

CO-CHAIR ROSENZWEIG: You mean the microalbumin test specifically.

MEMBER TAYLOR: Yes.

(Simultaneous speaking.)

MEMBER TAYLOR: The EGFR is not

CO-CHAIR ROSENZWEIG: Is the measure reliable?

MEMBER TAYLOR: We had evidence for reliability of the measure, too.

MEMBER MILLER: Yeah, the data had very high reliability for health plans, but it was a little less reliable for the physicians.

CO-CHAIR ROSENZWEIG: Oh, and less for physicians. Thank you.

MEMBER MILLER: Yes.

CO-CHAIR ROSENZWEIG: Any other comments?

(No response.)

CO-CHAIR ROSENZWEIG: So, let's

vote on this one.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. We have high, ten. Moderate, eight. Low, two.

CO-CHAIR ROSENZWEIG: Validity. Are the specifications consistent with the appropriate evidence? Exclusions appropriate? This is where Bob Bailey's issue may come up as a potential

MEMBER TAYLOR: Yes, about the EGFR. We didn't talk about that in the Subcommittee.

CO-CHAIR ROSENZWEIG: Yeah. Any other comments, Bob?

Yes, Vicky.

MEMBER DUCWORTH: This is just, I guess, kind of my complaining in general. Again, how I use metrics to evaluate any program really depends on the different criteria within these metrics.

And this is just one of those

metrics where it's really difficult for me to tie it to any particular or specific activity, because there are so many activities in it.

And it was specifically the ACE and ARBs that used to drive me crazy. And, again, I know because it's more indicative of a patient who was being treated for hypertension versus maybe nephropathy.

So, our providers would get credit for something they weren't, in fact, doing adequately. So, yeah.

CO-CHAIR ROSENZWEIG: I can make a comment on that. Just most physicians prescribe an ACE or an ARB to a diabetic because it is it's a positive impact on the kidney.

So, it's not that they're doing it by accident and luck, but they are deliberating choosing that agent because of its beneficial kidney effects.

CO-CHAIR GOLDEN: One issue, though, is that a lot of physicians will put

people on ACE and ARBs without measuring a microalbumin at all, which that's a subject of debate in various among different guidelines.

Yes, Sue.

MEMBER KIRKMAN: But I guess, you know, if we want performance measures to drive better outcomes, you know, I mean, you think about what's going to prevent end-stage renal disease from diabetes. It's, you know, looking for it or getting people on the right therapy.

So, ACE and ARBs, blood pressure control, which we have measures for, glucose control, which we have measures for. So, I mean, that doesn't really bother me that that is sort of part of the because that's really part of the downstream what you would do if you found microalbuminuria anyway.

So, I don't think it means the care was bad or that it was accidental. It means, for whatever reason, they're getting

what they would need anyway if the tests were positive.

So, whether the test was done doesn't matter so much. So, I think it's okay although it might drive you crazy from an analytical perspective. I mean, I think more performance measures should be like that where we're actually measuring that the right thing was done as opposed to the test was ordered.

CO-CHAIR ROSENZWEIG: Yeah, and once the test is positive, frequency of how often it needs to be done after that is a subject of great debate, because it can be used to help titrate up the medications you're using to treat it, but you don't necessarily have to continue to measure microalbumins forever afterwards once you've got them on the maximum dose of ACE or ARB.

Yes.

MEMBER HAYDON-GREATTING: What do you do if you have a population that can't be on an ACE or an ARB and, I mean, there's a

certain there's a percentage of growing African Americans that are having issues with the ACEs and ARBs. And then the new guidelines came out with calcium channel blockers and

PARTICIPANT: I think there are exclusions, I mean, if I remember correctly.

MEMBER HAYDON-GREATTING: I mean, this wouldn't exclude them, because you with those patients you would be making sure you had microalbumin, you know, tests, but

CO-CHAIR ROSENZWEIG: For treatment of hypertension certainly you're absolutely right. And, in fact, the new JNC 8 guidelines

MEMBER HAYDON-GREATTING: Right.

CO-CHAIR ROSENZWEIG: specify going to other agents other than ACE and ARBs for African Americans.

MEMBER HAYDON-GREATTING: Right.

CO-CHAIR ROSENZWEIG: So, you're absolutely right with respect to that. But

for people with diabetes, they're still recommending ACE or ARBs as first-line drugs.

MEMBER HAYDON-GREATTING: So, I just did a six-year longitudinal study on my employer group. And my patients down in North Carolina are I'm getting a small population of African Americans that cannot be on an ACE or an ARB because they've had some sort of reaction to it.

So, it's just a pattern I'm watching.

CO-CHAIR ROSENZWEIG: And there are a large percentage of our patients who can't

MEMBER HAYDON-GREATTING: And I think we didn't know that before. And now we're doing such a good job of putting people on them that have diabetes, I think we're starting to see more patients having some of those adverse effects

CO-CHAIR ROSENZWEIG: No, your point is --

MEMBER HAYDON-GREATTING: where it used to be rare.

CO-CHAIR ROSENZWEIG: Your point is well-taken. Some people will get cough on an ACE and they'll go to an ARB, and then they'll have elevated potassiums or

MEMBER HAYDON-GREATTING: Yes.

CO-CHAIR ROSENZWEIG: or they'll have an elevated creatinine

MEMBER HAYDON-GREATTING: Yes.

CO-CHAIR ROSENZWEIG: and you'll have to stop that.

MEMBER HAYDON-GREATTING: Yes, that's it.

CO-CHAIR ROSENZWEIG: Your point is well-taken, but usually those patients end up being referred to a nephrologist so that it would be considered

(Speaking off mic.)

CO-CHAIR ROSENZWEIG: Let's vote, yes.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. We have high, ten. Moderate, nine. Low, one.

CO-CHAIR ROSENZWEIG: All right. Okay. We're on feasibility. Sorry. Anyone want to speak to the feasibility the Workgroup want to speak to the feasibility aspect of this?

MEMBER TAYLOR: In our brief discussion of feasibility, we didn't see a problem.

CO-CHAIR ROSENZWEIG: The major issue might be is that you're collecting from different databases all at once for this one measure.

You can collect from with respect to billings for some, then medications for others, and then lab tests for the third. All of which could come from different origins, but I guess this has been in play for a while.

So, you've had no problems?

MR. REHM: I guess the question I have, I've run into some plans that don't have their pharmacy data available for mining, because it's in a PBM.

Is that an issue at all?

MEMBER TAYLOR: In my former life I used to work for AHIP, which is the trade association, and was in their clinical affairs. And now the PBMs are very much part and parcel of the data flow continuum.

MEMBER HAYDON-GREATTING: Also, the large employers are creating data warehouses where they're requiring all that data being in their back pocket now.

They're not depending on express groups or whoever else comes up to and part of their contracting, they're requiring an adherence clause so that they can get those numbers and look at that.

CO-CHAIR ROSENZWEIG: Any other comments?

(No response.)

CO-CHAIR ROSENZWEIG: Okay. Let's vote on feasibility.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: The results are high, 13. Moderate, seven.

CO-CHAIR ROSENZWEIG: Accountability, transparency, progress with respect to improvement. Do you have any data related to improvement over the last few years?

MR. REHM: Again on Section 1(b)(2) and importance is the performance data gaps between let's see. This is the tenth and the 90th percentile are about 13 percent in commercial. 13 to 15. 18 percent in Medicaid, and nine in Medicare.

And in terms of improvement on the mean, fairly stable with some improvement in PPO for commercial and in Medicare.

CO-CHAIR ROSENZWEIG: Okay. Thank you. Any comments by the Workgroup?

Bob?

(No response.)

CO-CHAIR ROSENZWEIG: Okay. Let's vote on this.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Let's all just push the vote one more time. We're missing two. Thank you.

(Pause.)

MS. BAL: Okay. We have high, 13. Moderate, six.

CO-CHAIR ROSENZWEIG: Okay. Let's vote on the overall measure.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Yes, 19.

CO-CHAIR ROSENZWEIG: So, NCQA should be happy with that. Okay. So, now we're going to move on to 2417.

DR. BURSTIN: And we'll come back to the foot measures. The Joint Commission

folks had to fly out of town. So, we're going to do the Joint Commission measures and then come back to

MS. TIGHE: Yeah, we're going to start with 2417. The way their set was developed it makes more sense to start there. So, apologies for the additional confusion on top of changing the agenda around.

(Pause.)

CO-CHAIR ROSENZWEIG: Okay. So, the title of this measure is "Risk Assessment/Treatment After Fracture." And the measure developer is the Joint Commission. And would you like to describe the measure for us, please?

And what's your name?

MS. DOMZLSKI: Cathy Domzlski from the Joint Commission. Hello, everyone. With me today is Ann Watt from the Joint Commission, and Dr. Ethel Siris who is the chairperson of our advisory panel.

We have had these measures in

development for the last eight years and, unfortunately, not much has changed over those eight years in the care of the osteoporosis and fragility fracture patient.

Our objective in measure development is that when used together although these measures are not paired, it gives an overall picture of the care for fragility fracture patients and our approach has several steps.

It begins with a literature review and formation of that advisory panel of experts, and they advise us at every step of the process.

We issue a call for measures. We develop a framework and draft measures. And those measures when developed, are then alpha tested for phase validity.

We invite public comment on the measures, and we then draft and specify the final version of measures.

We pilot test them, we do data

collection, and we do that at volunteer hospitals throughout the country. The results are subsequently analyzed.

The advisory panel reviews those results, finalizes measures. And in this case, we have three measures to present to you today relative to the fragility fracture patient.

The first numerical measure, Number 2416, encompasses lab testing for underlying causes of low bone mass.

Given the prevalence of secondary causes of osteoporosis, we feel this testing is essential to identify and prevent further bone loss and further fracture.

In lieu of testing for Vitamin D levels in that measure, oral D may be given during the hospitalization.

The second measure which is now the first one we'll discuss, Number 2417, seeks to ensure that those hospitalized with a fragility fracture indeed receive testing or

treatment for osteoporosis either while hospitalized, soon after discharge, or through the auspices of a fracture liaison service.

The last measure, Number 2418, addresses patients seen in the ED and sent home with a fragility fracture.

Now, there is no reason that those patients who are sent home should be treated differently or have a different standard of care than those patients hospitalized. And so, again, we want to ensure that bone mineral density testing is performed by treatment through a fracture liaison service or by referral to a testing facility or other practitioner. And that referral would be contained in the discharge instructions to the patient or caregiver.

Now, the results of our testing indicate in all three measure cases, compliance with the measures is at a mean level below ten percent, which is a little bit lower than that reported in the literature.

Generally speaking, 20 to 22 percent of patients are tested or treated for osteoporosis after a fracture, but certainly there's a lot of room for improvement.

We did find a ray of hope. There was one hospital in our pilot test group who enacted a couple of changes to their policies and procedures and they were able to achieve more than a 90 percent compliance rate.

And what did they do? They educated their house staff, their ED staff and they slightly modified their fracture order set and their discharge instructions to the patient. So, we feel this represents an opportunity to really improve care for these people.

So, once again we thank you for having us. We thank you for inviting us onto your previous call. Thank you very much.

CO-CHAIR ROSENZWEIG: Thank you. Thank you very much. Would the Workgroup members like to comment on that?

MEMBER KEARNS: Okay. So, I was the lead on this measure, the first one we're going to review, which is the assessment for fracture risk or treatment by several measures in patients who are actually hospitalized with a fracture. So, not the emergency room. This is inpatients who are dismissed and the measure to meet the measure so, it's every patient over age 50 from an inpatient with a long list of potential fractures that would meet this diagnosis.

And to meet the measure, the numerator would be had a DXA scan ordered or performed, prescribed a medication or who were seen by a fracture liaison or had some other risk assessment measured if DXA was not performed.

There were a couple of exclusions. People younger than 50, people who are already on a treatment for osteoporosis or enrolled in an osteoporosis trial, patients who are on comfort measures only or those who had

expired, or those who had documented bone mineral density test in the last 12 months prior to the fracture.

So, that's the basics of the measure. I could go right to the evidence if that's okay.

I think the Working Group agreed that the evidence for detection and treating osteoporosis with the ultimate goal of preventing additional fractures was very high, that there's no reason to question that medications are helpful.

Whether there is evidence that scheduling a DXA is the same as prescribing a medication, I think, is not clear, but certainly treatment of osteoporosis goes up if there's more DXAs performed.

So, it may seem to be an intermediate measure that's reasonable. So, that's the evidence I think any questions? There's other concerns and things that will come up later about the process and things,

but I think the evidence for treating osteoporosis, especially in patients who had a low-energy fracture, is pretty solid.

CO-CHAIR ROSENZWEIG: It specifies specific fractures that are supposed to be fragility fractures, but what if you have fractures that are technically not fragility fractures that would have those same diagnosis codes?

For instance, if a person has existing prostate cancer or breast cancer with metastasis to the spine.

MEMBER KEARNS: That's a good question. I looked at there's a lot of codes in there. So, I'm not an expert in all the codes.

And I don't think that it included things that were considered pathologic from cancer treatment from my looking through there. But, again, this may be people who are more familiar with codes could speak to that, but I don't think it includes that.

So, I don't think to answer your question, I don't think that patient would be captured by this, but there are a lot of codes included.

CO-CHAIR ROSENZWEIG: Okay.

Yes.

DR. SIRIS: Well, the intent was that these would be osteoporosis-associated fractures. And by definition if you have metastasis in the bone, it is not an osteoporosis-associated fracture.

CO-CHAIR ROSENZWEIG: But you're making you're using the DXA test in order to diagnose osteoporosis and it's the fracture that comes first, isn't it?

DR. SIRIS: No, no, no. We're trying the diagnosis of osteoporosis can be made based upon a DXA that gives you a certain level of lowness. Or if you have low bone mass and you've had one of these fractures, clinically you have osteoporosis.

And since these are generally

fractures that are operated on in the hospital, you would exclude metastasis, because that would be noted. Or if you were a patient who broke a hip in the setting of being treated for prostate or breast cancer, clearly the diagnosis would have to be made before you could call it an osteoporosis-associated fracture.

That is to say you'd have to exclude as part of your clinical care that there was metastatic disease in bone. I mean, that's what would happen.

We are not interested in pursuing fractures once the diagnosis has been made that it's due to metastatic disease in bone. That's a separate issue.

We're trying to capture the patient with a fragility fracture due to osteoporosis, which clinically means you either have a score on DXA that's minus 2.5 or below with the spine or hip, or you have low bone mass and you've broken one of the major

bones that are considered osteoporosis-associated fractures.

Right now those folks get the fracture fixed by the orthopedic surgeon and then they're told goodbye and good luck. And there's no further effort to reduce the risk of the next fracture for which they are at very high risk, and that's what we're trying to get past.

CO-CHAIR ROSENZWEIG: So, to follow up on this question, there's an Excel spreadsheet attached which we --

DR. SIRIS: Yes.

CO-CHAIR ROSENZWEIG: So, the metastatic or the pathologic fractures would be in the spreadsheet?

DR. SIRIS: They're excluded. Sorry. Okay. Short answer, they're excluded.

PARTICIPANT: Okay. I'm sorry. I may have misinterpreted the way it was set up, but I assumed that it was the fracture that was the initiating event that put the person

in, you know.

DR. SIRIS: The reason for the fracture is osteoporosis. And right now that's not being recognized.

So, these are people who have had the complication of osteoporosis. They've had a fracture. Generally speaking, they've never been diagnosed as having osteoporosis. And, therefore, no attempt is made to treat them to prevent the next fracture for which they're at high risk, but these are for patients admitted with a fragility fracture. That's correct.

MEMBER KIRKMAN: So, I guess the, I mean, just to follow up on that, I guess the only question is, is it specific enough in the measure specifications that's redundant that those people wouldn't be, I mean, I think we all get it that that's not what we're looking for.

But if you're measuring this, is it specific enough? Because I don't see an exclusion for metastatic cancer or

CO-CHAIR ROSENZWEIG: I think that that gets into specs, and not into the evidence.

MEMBER KIRKMAN: Yes, sorry. And ICD-10 will fix it anyway.

CO-CHAIR ROSENZWEIG: Okay. Any other comments?

Yes, Bill.

MEMBER TAYLOR: We're asking question about the evidence?

CO-CHAIR ROSENZWEIG: Yes.

MEMBER TAYLOR: So, could someone, an expert on this, comment on why a T-score of minus 2.5 is the cutoff for years?

DR. SIRIS: It's an arbitrary cutoff that the World Health Organization came up with in 1994 when a bunch of osteoporosis experts got together in Geneva, and over beer and pizza they decided that the cut-point could be minus two, or it could be minus three, and they finally concluded that it should be minus 2.5.

It was arbitrary. It's not unreasonable. The lower your T-score, the greater your relative risk for fracture.

And this became an operational definition of osteoporosis in 1994 from the World Health Organization as a way of helping epidemiologically to sort out those people at higher risk of fracture and it's what we live with.

We have just published a paper in Osteoporosis International calling for an expansion, that is to say a consensus group is saying that that's one good way to identify the patient at high risk.

Osteoporosis is a disorder of reduced bone strength that predisposes to a high risk for fracture. The T-score is one way to do it.

Another way to do it is you say to somebody, you just had a hip fracture and you're 75, you have osteoporosis regardless of T-score.

If you have low bone mass not quite at minus 2.5 and you've had certain types of fractures, that puts you statistically at very high risk of another fracture and that's been well-established in the literature. So, that's another way to make the diagnosis.

And I think you're right that the ICD-10 codes are going to help us there, but that's the answer to your question as to how they picked minus 2.5.

CO-CHAIR ROSENZWEIG: I think if it was the World Health Organization, it wasn't a meeting over beer and pizza. It was probably wine and quiche or something like that.

DR. SIRIS: I know the people who were there and I believe that was part of it. And the reason it's called a T-score is because the guy from one of the bone density companies was named Tom, and they decided to name it for him.

They were drunk when they did this.

(Laughter.)

CO-CHAIR GOLDEN: So, can I follow up on your comment just so because one of the reasons for the cut-point which actually gets to be interesting because you get into overuse, the data on using the drugs, the phosphonates in people under 2.5 because they are toxic in their own way

DR. SIRIS: I would argue that.

CO-CHAIR GOLDEN: Well, I mean, I've been getting different kind of depending on who you talk to, but I'm just saying I've seen actually people getting more treatment below 2.5. And I'm not sure well, at least from my perspective in

DR. SIRIS: The literature shows that a number of the clinical trials had entry criteria that included people with hip T-scores that were the basis for entry and many of those were below minus 1.6.

Some of the pivotal trials included people with osteopenia who had already had a vertebral fracture. So, there's quite a bit of data indicating that these drugs do work.

Risedronate studies have shown that the drug worked in people who were misclassified as osteoporosis who actually had osteopenia. And I think clinically today if you get somebody with a bad fragility fracture and the T-score is minus 2.3 instead of minus 2.5, you're going to treat.

We also use algorithms like FRAX, a fracture risk analysis, which doesn't necessarily require a bone density although it works better if you do a bone density. And it will show that if you're an older individual, 70, 75, 80 and you're osteopenic and you have one other risk factor such as the fracture you just sustained, your risk is going to be very high.

And the current guidelines

recommend that those patients should be treated to lower the risk of the next fracture and there's a literature that suggests for several drugs that it does reduce fracture risk in those patients.

CO-CHAIR GOLDEN: So, when you refer to a fracture risk assessment, you're specifically referring to FRAX?

DR. SIRIS: Well, FRAX would be the common one used in the United States. There are other algorithms that some people choose to use, but FRAX is the one that is sort of WHO-II. Bone density was WHO-I.

CO-CHAIR GOLDEN: Okay. Thank you.

MEMBER TAYLOR: So, just to clarify, I understand that as T-score goes down greater values below fracture risk goes up. And as FRAX score goes up, fracture risk goes up.

But the trials that show bisphosphonates prevent fractures, haven't they mostly been limited to people who not

only have low bone density, but also have had a fracture?

DR. SIRIS: Some of the trials used exclusively cut-points of minus 2.5 at spine or hip as entry criteria. Other trials including those in post-hip fracture patients with the drug zoledronate, included people who had had the hip fracture and they simply had to be osteopenic.

I mean, they did not require a minus 2.5, and the drug was highly effective at reducing the risk of subsequent fractures and reducing mortality.

MEMBER TAYLOR: But for people with only the low bone density less than where they did not achieve a greater than 2.5 reduction and they had no fractures, those people have not been studied and shown to

DR. SIRIS: No, they have, because some of the trials back when they were first enrolling people, there were differences in the two manufacturers' reference populations

and it turned out in retrospect that a fair number of people whose T-score was minus 1.6 to minus 2.4 were enrolled. And in the alendronate trials, there was an effective reduction in risk of certain fracture types in that setting.

Now, of course, whether or not the drugs work in people with osteopenia who haven't had fractures is a good question, but every patient that we're talking about has had a fracture.

So, it's a different group of people where the risk is much higher and where we have evidence from a number of trials, particularly the zoledronate trials in hip fracture patients, that the drugs are highly effective and that you shouldn't get too hung up as to whether the T-score was minus 2.2 or minus 2.5 or minus 2.7.

MEMBER KEARNS: Well, I just wanted to say that I agree with the intent of the measure.

If you pull up the appendix that lists the other ways for a fracture risk assessment, if you could pull that up, because that is not clear to me, the measure as written talks about DXA.

Maybe it could be modified to say bone density for which there are several ways to assess it. I don't know, but the ancillary information is not clear what those are in the appendix.

If they're not in the appendix, then they should - since they are a way to qualify for the measure, they have to be very specified.

And what I had included, an ultrasound of the humerus or something, which I don't - when I open it on my computer, that's what it looks like.

MS. DOMZLSKI: Our first submission was --

MEMBER KEARNS: It's in the appendix.

MS. DOMZLSKI: - in ICD-9.

MEMBER KEARNS: It's in the appendix.

MS. DOMZLSKI: In the translation to ICD-10, there are additional codes that need to be supplied. But the other assessment methods that may be used are the QCT of the spine, the QUS of the heel, DXA of the forearm, a SXA or DXA of the heel, and the FRAX assessment, too.

And we are working on supplying the current, up-to-date codes for those.

MEMBER KEARNS: Yes, those sound like the right ones. Just the file that we have doesn't say that clearly.

CO-CHAIR GOLDEN: And just to help me again on the evidence, I know that DXA scans are very good. All of us have seen people coming in with the local heel scans in somebody's office. So, are those considered valid and useful by the evidence?

DR. SIRIS: There's evidence that

if you're very low on one of these peripheral tools, your fracture risk is elevated.

Clearly the preferred test is DXA. DXA right now is a relatively inexpensive test and there are a fair number of DXA machines around. So, most people are likely to get a DXA.

But if you lived in a place where the nearest DXA facility was three hours away and your physician, after you had your shoulder fracture, did an SXA which showed low values, that would satisfy the measure.

MS. WATT: Can I just clarify? I'm not sure that it's clear. This measure really is about the denominator is patients who come in with a fracture, and the numerator is patients who had either a DXA ordered or performed, or a prescription for medication while they're in the hospital, or seen by a fracture liaison service.

That is simply it in terms of what this measure comprises.

MEMBER TAYLOR: Well, that's not what she just said.

(Simultaneous speaking)

MS. WATT: Sorry.

DR. SIRIS: There are other fracture risk assessment tools which are very unlikely to be used, because it's easier to get these other things done.

Remember this is - because this is potentially a Joint Commission measure, what's in most of our hearts is the hope that right now probably the best way to get the post-fracture patient treated for osteoporosis is through the fracture liaison service. The evidence for that is powerful. Kaiser has shown it really, really works.

The problem is that most hospitals are unwilling to pay the salary of a fracture liaison person.

And if this were to become a Joint Commission measure which will require your approval before they will approve it, it would

conceivably be an incentive to hospitals admitting patients with fractures - I'm telling you something that isn't in the document you're reading, but the fracture liaison service is a powerful way to get these people managed.

It's a coordination of care effort that puts the fracture fixers together with the subsequent fracture preventers. It works. There's a plethora of data.

And yet, right now people are struggling to put these things in place because of the cost of hiring somebody for 70,000 bucks a year to do the work.

So, the hope is with something like a Joint Commission measure, that it may be an incentive to a hospital to spend the money to get fracture coordinators and make it work the way it does at Kaiser.

CO-CHAIR GOLDEN: Okay. I guess -

DR. SIRIS: No, no, no. They don't have to do that. They can make sure that as

that hospital you visited does, that they know they should do it and the endocrinologist agrees that they'll see everybody who's had a fracture.

CO-CHAIR GOLDEN: One more technical question.

DR. SIRIS: Yes, sir.

CO-CHAIR GOLDEN: These activities, this assessment has to be done by discharge, or within 60 days of discharge? Is there any framework that it has to be before discharge?

MEMBER KEARNS: All these measures to meet the criteria have to be by - according as written as the time of dismissal, but that includes just an appointment for or just an appointment with a fracture liaison.

So, you don't actually have to do the DXA which comes down to a feasibility issue, not really an evidence issue.

CO-CHAIR ROSENZWEIG: What about if the patient has already had this evaluation prior to the fracture like before coming into

the hospital?

Do you have like a time frame of -

MEMBER KEARNS: As written, it says if they have a documented DXA within the 12 months previous.

CO-CHAIR ROSENZWEIG: Previous, okay.

MEMBER KEARNS: Now, there are feasibility issues with that, again, but the evidence is what we're discussing, I think.

CO-CHAIR GOLDEN: We've covered multiple fronts here. We've covered some specification issues which will help us later. We have a point at the end there.

MEMBER LEE: What is the definition that you're using a fragility fracture? Can you just clarify that?

CO-CHAIR GOLDEN: There is a spreadsheet, apparently, that you can look at and it's probably on the SharePoint. That lists everything that you - probably more

than you'd want to look at.

DR. SIRIS: I'm looking at that right now. I think it's in front of me. It includes pathologic fracture of -- the Excel column cuts off, but it gets back to our question of just how does the measure exclude metastatic fractures. It's just not clear from this.

DR. SIRIS: Let me start by saying that some orthopedic surgeons if they treat an osteoporotic fracture, call it a pathological fracture because they get paid a little more if they call it that, but the intent is that these are osteoporosis-based fragility fractures not due to things like metastatic cancer.

Cathy.

MS. DOMZLSKI: The patient with pathologic fracture as Dr. Siris indicated, is very often not due to a metastatic lesion.

However, there are also cancer patients who the first sign is a fracture that

they have an underlying disorder that's of a malignant nature. And so, that doesn't mean that they should not be tested or treated.

Patients who are far advanced in cancer who are on comfort measures only are excluded from the measure.

CO-CHAIR GOLDEN: Okay. Let's get back to evidence. So, we have wandered around here a little bit and are we ready to talk about or vote on the evidence of this measure?

MS. BAL: Voting is open.

We have high nine, moderate ten.

MEMBER KEARNS: Okay. If we move on to the performance gap, I think that the published literature would support that there's a huge performance gap.

And they did do pilot studies in some hospitals and shown that there was quite a bit of a gap in the care here. So, I would rate that this is a high.

We did not in our working group, get a lot of time to discuss this measure,

because we started with a different one. And, as you've already seen today, sometimes it's hard to get to all the measures, but I think there was a consensus that this was a high gap.

CO-CHAIR ROSENZWEIG: Ready to vote.

MS. BAL: The voting is open.

Okay. We have high 17, moderate two.

MEMBER KEARNS: Okay. And then we move on to priority, and I think that, again, if I can speak for the working group and if anyone wants to jump in and correct me, that we felt this was a high priority that this is probably an area that's been neglected in terms of performance measures and that the impact on the health of individuals and society was great enough that make this a high priority.

CO-CHAIR GOLDEN: Seeing no comments, are we ready to vote?

MS. BAL: Voting is open.

The results are high 18, moderate one.

MEMBER KEARNS: Okay, the reliability measure. They did do testing in some hospitals and about reporting. And it's appeared to be quite a reliable way to ascertain the data.

CO-CHAIR GOLDEN: It gets into the issue about your specificity, I mean, how well you identify your denominator and your numerator.

So, any other further comments on this issue?

MEMBER KIRKMAN: So, am I right that there were some things that you did not do reliability testing on like the fracture liaison service and whether they were already on osteoporosis treatment and whether they had had a DXA before, the year before?

MS. DOMZLSKI: We tested, for reliability, we tested every data element.

MEMBER KIRKMAN: Okay.

MS. DOMZLSKI: Whether they had been tested in the previous year, everything that you see listed is a data element with reliability tests.

MEMBER KIRKMAN: Okay. I guess it was a staff comment that maybe it was just a mistake.

MS. WATT: Yes, if I understood your question correctly, it was did we see how many people were referred to the fracture liaison service versus had the DXA versus were on the medication?

The answer is no, because that's one data element. We didn't look for the individual component of that data element. We just looked to see if one of those was done.

MEMBER McDERMOTT: But I would ask do -- some of these people potentially are a refracture. It's not a first fracture or they are known osteoporosis.

If they are known osteoporosis -

MEMBER KEARNS: They're excluded. If they're on an osteoporosis medication already, they're excluded.

MEMBER McDERMOTT: Oh, I missed that. Thank you.

MEMBER KEARNS: And if they had a DXA within the last 12 months, they're excluded. So, it would only include, hopefully, people who were not diagnosed or treated.

Now, people can be diagnosed and not be treated, or been given a prescription and not taking it.

MEMBER McDERMOTT: Thank you.

MEMBER TAYLOR: What's the test-test reliability of the DXA?

DR. SIRIS: I'm not sure I know what you mean by test-test reliability. DXA is a pretty precise tool. I mean, it's an accurate and easily done test.

MEMBER TAYLOR: Yes, if you do it twice, do you get the same measure? That's

what I mean.

DR. SIRIS: Because of the way it's calibrated, because of the way the machine works if you're trained and you know what you're doing, you should get very, very, very tight similarity in repeating the test twice, yes.

MEMBER TAYLOR: And then when it's looked at and actually used, do people actually well-trained and know what they're doing, I mean, when a physician gets -

DR. SIRIS: Well, I mean, a lot of them are in radiology practices today. And the radiologists seem to know what they're doing.

The ones that are in endocrine offices or primary care offices, you hope that they are giving you decent data.

MEMBER TAYLOR: Oh.

DR. SIRIS: It's not perfect, but it's pretty good.

MEMBER TAYLOR: And how about at

different sites, you know? The spine and the total hip and -

DR. SIRIS: Generally what we do is we measure the spine and the hip and you can certainly get discordance between spine and hip, but that's part of the natural history of the disease. That's not the machine.

Some people will be low at the spine and better at the hip. Other people as they get older and older and get more degenerative disk disease, will have spines that can't really be measured, but the hip can still be very well measured correctly.

And many places will include a forearm measurement which takes another 30 seconds. If the spine is useless, you can look at the forearm and the hip.

MEMBER TAYLOR: And how about in the evidence base? Were those numbers in wrist, forearm, spine used when you gave us the T-scores that have been used to show fracture reduction with treatment?

DR. SIRIS: Well, no. In terms of making a diagnosis, the spine, the hip, and the forearm were the standard sites. And those are the standard sites that are still used today.

For clinical trials, entry level usually involved looking at spine and hip and you had to be low at one or the other.

If you're low at any of those two sites where there were two hip sites and a spine site, if you're low at any of those three sites at the level of minus 2.5 or below, you were called osteoporosis.

CO-CHAIR GOLDEN: So, again, Sue, do you have something you want to, or are you just

MEMBER KIRKMAN: Well, I'm still confused about your reliability testing. Because at least in 2(a) 2.3 it looks like you looked at what happened in the hospital, but not whether they came in on FDA-approved pharmacotherapy, whether they had previously

had a DXA in the prior year or the fracture liaison service.

Am I misreading that? Because, I mean, that's the same thing that the staff comment said that you didn't do reliability testing on all the parts of the numerator.

MS. WATT: Those are exclusions to the denominator. They never would have made it into the measure.

MEMBER KIRKMAN: Okay. But you can reliably pick out those exclusions, because - okay. I mean, because otherwise a lot of people are going to score poorly on this even though the patient was already on appropriate therapy. You know what I'm saying?

MS. WATT: I do know what you're saying, but the thing is those patients - there's the whole mass of fracture patients, and then we look for those exclusions before they ever even get into the measure.

We're talking about the reliability for this measure. Those patients

aren't in there.

MEMBER KIRKMAN: I mean, don't you have to have a reliable way to find the exclusions?

Maybe the NQF staff can help me here, but - so, we have to kind of take your word for it that you could reliably exclude people.

MS. DOMZLSKI: For example, every one of those exclusions, for example, a prior diagnosis of osteoporosis, becomes a data element when you collect this data. And we did reliability scores on all of the data elements.

In other words, how accurate was the ability to identify that particular data element?

And for that particular data element, the kappa score was 0.75, which is quite high.

We had a match rate of almost 95 percent in terms of what we abstracted

compared with what the hospital had abstracted and identified as a patient with a prior diagnosis of osteoporosis.

So, that was one data element that was -

MEMBER KEARNS: So, maybe, Sue, you're asking about, I mean, there's some different questions that I have within this, but they're not part of the reliability. And one of them is it's very difficult to know whether a DXA has been done at a different center within the last year, but that's a different question in my mind than what the hospital is looking at and what you are confirming were very well-matched. And that's, to me, the reliability.

The feasibility and usability are different things about certain measures, but the reliability I think they did test and was found to be quite good with the limitations that the hospital may not know everything because of the nature of healthcare in this

area.

CO-CHAIR GOLDEN: Patricia, do you have something or are you just vestigial there? Okay.

MEMBER MILLER: Does Sam's Club do DXA scans?

(Laughter.)

MEMBER KEARNS: Not yet.

PARTICIPANT: Costco, I'm sure.

DR. SIRIS: DXA scans involve radiation so that they have to be performed in places where you have - in most states you have to have licensed x-ray technicians. And they do involve radiation and they're not - now, I suppose Sam's Club may do ultrasounds and things like that, but it's never been profitable. So, I doubt that it's being done.

CO-CHAIR GOLDEN: Are we ready to talk about reliability and vote?

MS. BAL: Voting is open.

High eight, moderate 11.

MEMBER KEARNS: Okay. So, the next

point to talk about is validity and that is the strength of the evidence of the different items in the measure and whether that would be a valid assessment of quality.

And I think, again, our working group did not get a chance to really discuss this in this level of detail, but I think we would all agree that assessing for osteoporosis by one of these measures after a fragility fracture would be a valid assessment of quality.

CO-CHAIR GOLDEN: Ready to vote?

MS. BAL: Voting is open.

High nine, moderate 11.

MEMBER KEARNS: Okay. I think the next point to discuss is feasibility and I think there is some points to discuss here potentially.

This is at a facility level where we're talking about inpatients. So, I think that's important to keep in mind because of the way that healthcare is structured around

inpatients. A lot of the tests we're talking about are not routinely done as inpatients.

That's why I think allowing for some of the other measures and/or an appointment for a measure gets around that because anybody who knows DXAs, knows you're not getting a DXA as an inpatient. In my facility you can't, because the machines aren't in the hospital.

One could argue about whether an appointment is the same as a measurement of it. Because in my experience with hip fracture patients when we tried that giving them an appointment, maybe ten percent of people would show up for the appointment.

So, you know, that doesn't negate the importance of doing it and having a measure and certainly some populations might require a different approach. And I think the fracture liaison service nicely gets around a lot of that.

So, I think, again, maybe perfect

shouldn't be the enemy of good here.

DR. SIRIS: I think you're correct. The fracture liaison service is one way to get around it.

The other way to get around it is to discharge the patient on osteoporosis medication if you choose to do that.

MEMBER KEARNS: Our surgeons do not like that.

DR. SIRIS: I understand. But, I mean, that is still an alternative. And many surgeons have now determined that you can safely - there's no evidence in the literature that you delay fracture healing, but you're right. There are biases.

But the point is it can be - you have three mechanisms. One of which is to schedule a DXA. One of which is to provide treatment. And one of which is if you have an FLS, that's the coordination of care mechanism that would work best.

And that's what we're sort of

hoping if this measure goes forward, will become easier to do.

CO-CHAIR GOLDEN: A question for you, Ann. Do you have a perception of what the burden of data collection is for this?

MEMBER KEARNS: Well, I think that's another point because there is very hard in - and so, within a system it's easy. Maybe with an electronic record, that would include DXA in it. Not all DXA machines are included in electronic records even in my expanded healthcare system.

So, finding out if they've had one in the last 12 months could be tricky, could lead to some duplication, which is not well thought of by most of us because that's an unnecessary thing.

But, again, if you have an appointment and you're liaising with a primary care physician who might know that, you'd like to think it wouldn't get duplicated, but I think it's a valid point that there is the

risk that there could be duplication.

Relying on patients to know what a DXA is, a bone scan or an x-ray is not always so reliable. I think that's a low risk, but real.

DR. SIRIS: Could I just comment that if you had a DXA last year and this year you broke your shoulder, it wouldn't be such a bad thing to have another DXA if it turned out that they goofed and they couldn't find out you had had one before, I mean, clinically.

CO-CHAIR GOLDEN: I think we're ready to vote.

DR. SIRIS: Yes.

MS. BAL: Voting is open.

High two, moderate 11, low six.

MEMBER KEARNS: Okay. And the next point is the usability and use. And I don't think - we don't have any - since this is a new measure, there's no prior usability other than the pilot studies that were done that

showed actually that at least in one system they were able to really step up to the mark.

But I think there is definitely published data from other systems that this is a very usable system.

I think, again, the same caveats about DXA and inpatients and those things probably apply, but I think this would be rated at least a moderate, if not high.

CO-CHAIR ROSENZWEIG: Since this is a Joint Commission measure, is the accountability at the level of the hospital, or at the level of the providing physician?

MS. WATT: These data would be aggregated at the hospital level, not at an individual provider level.

CO-CHAIR ROSENZWEIG: Okay. So, if a patient had the DXA performed as an outpatient shortly after the admission, that would be okay, I assume.

MEMBER KEARNS: If they had the appointment at the time of dismissal.

CO-CHAIR ROSENZWEIG: The appointment, okay.

MEMBER KEARNS: They would meet the criteria. I think the limitation is that there are - and certainly I'm learning in my institution this is an inpatient measure. And there's an ER measure that talks about fractures, but there are a group of people that we're missing with both of these who don't - and those are especially the vertebral fracture patients who present to their physician, who get an x-ray with an incidental noting of it done for other things.

So, this is definitely the best we have. And I think making it a facility measure is important because we've shown it doesn't happen, but this won't address all patients, in my opinion.

CO-CHAIR GOLDEN: Going once.

Bill.

MEMBER CURRY: From the usability issue, I think it was mentioned that I think

the surgeons are going to have some reticence to be ordering tests that perhaps they think should be followed up by a patient's primary care provider.

I think they're going to have reticence about ordering medications that will be ongoing, prescribed maintenance-wise by their primary care provider.

And I think the usability of this will be difficult because of that unless there's a fracture liaison service.

DR. SIRIS: Which is precisely the point. In other words, it's very, very simple to reassure the orthopedic surgeons that others in the hospital are very happy to take on that responsibility. That's been shown to be highly effective, because this would be hospital policy.

The orthopedic surgeons are not obligated to do this. They're simply obligated to cooperate if somebody else is willing to do this, and that's worked.

The reason right now there is such a tremendous gap is precisely because the orthopedic surgeons fixed the fracture very well and that's it. This is an attempt to make sure that there's the continuation of care and the link that will make sure these patients get treated.

MEMBER CURRY: But as the measure stands today, I think it would be difficult to use this, because many facilities do not have the fracture liaison service.

DR. SIRIS: Right, but many facilities, hospital facilities have internists. And hospital facilities have endocrinologists.

And if the hospital recognizes that this is a critical quality care measure and that we have an epidemic of fractures, as we do, the hospitals may simply decide that an endocrinologist will be asked to see everybody who comes in with a fracture the same way many patients prior to surgery have to be seen by

an internist to be screened for surgery.

It's feasible. It's doable.

CO-CHAIR GOLDEN: Well, you're getting a little far afield.

DR. SIRIS: Sorry.

CO-CHAIR GOLDEN: And we're also talking about how to spend money. So, that's another interesting question as opposed to other things.

MEMBER BREEN: I'm just going to comment I think that's why it's important that it is a hospital-based measure, because right now the hospitals are recouping large sums of money on operating on the first fracture, the second fracture, the third fracture. It's true.

And I think the resources are potentially already there for some cost savings in the larger scheme of things, but the burden has to be on the hospital to coordinate the resources that many of them already have. They're just not linked, right?

So, I think you're right. The usability, it's not the easiest thing right now, but that doesn't mean that it's not the right thing.

CO-CHAIR GOLDEN: The hospital has its own ways of solving problems.

MEMBER KEARNS: And you would be surprised. The orthopedic surgeons have their own initiative called Own the Bone. So, they are more aware of it.

Although they don't want to take responsibility for doing it, in my experience they're very willing to let me come in and - so, I think there won't be as many barriers. We just need a little incentive.

CO-CHAIR GOLDEN: Bill, do you have a comment?

MEMBER TAYLOR: I got a question. Is there a concern with this measure that if all that's required is ordering the DXA as opposed to somehow making sure the patient gets the DXA or somehow connecting to the

source of care that would follow up appropriately, that there's any concern that this wouldn't get far enough to actually close the loop?

DR. SIRIS: No, I think right now the problem is nothing is done. And it may well be that if you go the first step, which is to even think about a DXA, you recognize that there's a clinical issue.

So, while ideally, you know, you want the DXA done and you want somebody to actually look at it, it's the first step toward moving into a paradigm where you're going to do the right thing, I would hope.

MEMBER BREEN: And I would just comment in order to order the DXA, you need to diagnose the patient with osteoporosis which is often not happening, right?

In order to order the DXA, you have to put down qualifying criteria. And that then enlarges your capture rate of these patients as opposed to patients just coming in

with "a fracture" who then leave without a diagnosis on some level of having osteoporosis.

So, I think the ordering, you know, in and of itself has some good things even if the test doesn't.

CO-CHAIR ROSENZWEIG: So, getting back to usability and use. So, do we have accountability, will it result in improvement and are benefits better than the risks?

MEMBER KEARNS: I would say yes.

CO-CHAIR ROSENZWEIG: Okay. Any other comments?

MS. BAL: Voting is open.

High seven, moderate ten, low two.

CO-CHAIR ROSENZWEIG: We're up to yes and no.

MEMBER KEARNS: I would think that we should vote yes, but I'm maybe a strong advocate for this overall.

MS. BAL: Voting is on.

Let's all try one more time.

We're missing one person.

We still need one more. We need 19.

So, 19 yes.

CO-CHAIR GOLDEN: Thank you. Nicely done. Are we going to that one now? Okay. We're going backwards.

MEMBER BREEN: So, I'm so glad we changed the order of this. Can I just say I was worried we were going to start here and then go - exactly, never finish.

That's what happened on our workgroup call. So, you know, I'm sorry. So, this measure is Laboratory Investigation for Secondary Causes of Fracture.

This is getting at the concept that many people who come in with osteoporosis actually have a secondary reason besides having senile osteoporosis or age-related osteoporosis.

And so, I'll get down into this. In terms of the background, there have been

studies that estimate that when you look at patients with fractures, a large percentage, anywhere from 40 to 50 percent, actually have some underlying cause that can be identified with laboratory testing that is not currently happening.

The numerator statement for this measure, patients who have all, actually, have assessed their laboratory tests ordered or performed prior to discharge. That's CBC, kidney function tests, liver function tests, a serum calcium, a 25 Vitamin D level or the provision of Vitamin D.

And that kind of gets back to what we were talking about in that kidney measure with whether you measure urinary microalbumin or whether you treat with an ACE or ARB and you get credit for both, right?

You get credit for the measurement, but then you also kind of get credit for just treating the assumption of the deficit.

The denominator, patients over the age of 50 who have been discharged from an inpatient status. Again, these are admitted inpatients with a fracture. So, it's a similar group as the one before.

And exclusion criteria, again, comfort measures only. That's important in this group that we're taking the amount of people who had been enrolled in clinical trial pertaining to osteoporosis if they had had laboratory testing performed in the prior 12 months.

So, let's talk about the data if anybody has any larger questions. I don't know if the developers want to talk, make a comment at all about this particular measure and the thought process behind this particular measure.

DR. SIRIS: Well, right now nobody gets treated or worked up. Twenty percent get treated or worked up. So, the assumption is that they're not getting the testing either.

I mean, we don't have evidence that they're getting the testing with the intention that it would be used to determine whether it related to their fracture.

You can't use medication until you know that the patient doesn't have some other medical condition either because the medication won't work or because you're going in the wrong direction. You've got to deal with that first. So, this is sort of a logical part of the process of assessing the patient.

MEMBER BREEN: So, for the rest of you, this is what our entire work group time was spent on this concept of how you parse measures that hadn't been looked at specifically.

So, when we get into the challenges around the evidence for this, there is little data that has looked specifically at what happens when you measure these laboratory tests, because nobody does that.

We have large kind of osteoporosis studies that use these laboratory assessments as part of their study, but nothing specifically looking at this measure.

But, again, you can't do the treatment that we just said we need to do without having these tests done. So, it's a little bit of the chicken and egg.

DR. SIRIS: Sometimes people do use the treatment and that's the wrong thing medically.

MEMBER BREEN: So, there was some just in terms of the data, there was one trial looking at kind of getting at this whole concept, you know, when you do all this testing and reading, you do a lot of interventions at the time of the clinical event, i.e., the hospitalization and you do the testing and you assess the fracture, there's a higher rate of patients getting treated down the road.

So, again, I think that's the best

supporting evidence to suggest that testing patients for secondary fractures in the hospital during the time of their fracture is supported by the data.

Anybody else have any -

MEMBER BAILEY: Just a quick question. Why is PTH not included? It may be just my ignorance in terms of not being close to this field.

MEMBER BREEN: As one of Dr. Siris' former fellows, I'm going to defer to her to comment on this.

DR. SIRIS: If you ask a dozen endocrinologists what should the list of blood tests be, you'll get a dozen answers.

So, what we were trying to do is to take, you know, what do you really need to know?

You need to know that the calcium is not 14. You need to know that the patient is not profoundly anemic, which could be suggestive of a number of other disorders.

You need to know that the patient is not in renal failure. And it's probably not a bad idea to also know that the patient doesn't have liver failure at the time you're seeing the patient with a fracture. So, we went with the bare minimum.

Now, once you've done that, you can probably start the patient on treatment. If you're missing normocalcemic hyperparathyroidism, the drug is still going to work. And hopefully because you're treating the patient for osteoporosis, there will be ongoing care that we'll be able to continue to evaluate the patient, but at a minimum.

Plus, most of these tests are things - maybe not the liver panel. Most of these tests would be done as pre-op tests with an eye toward not putting somebody in renal failure into the OR to have their fracture fixed, but it may not be thought about in terms of its role in playing, you know, what's

going to happen to the patient clinically with the osteoporosis.

So, it's a way of forcing these standard, not terribly expensive tests to set you up. Yes, there are other tests that could be done.

What about 24-hour urine calciums, you know, all kinds of things, sure.

MEMBER BREEN: And this is the challenge when you look at the data is that, you know -

DR. SIRIS: Celiac screens and -

MEMBER BREEN: You know, for every 12 bone studies, you have 12 different constellations and data pieces that they've done.

So, this particular constellation of data measurements has not been studied in and of itself as a standalone unit if we're looking for evidence. But my poor man's take on it was a less well-spoken version of Dr. Siris'.

What do you need to know to not kill somebody or hurt them really badly while they're in your care and you're trying to treat them for osteoporosis?

CO-CHAIR GOLDEN: So, what you're saying is we've covered most of the waterfront or the important waterfront aspects. May not be perfect, but it's good enough.

MEMBER BREEN: Yes.

CO-CHAIR GOLDEN: Okay.

MEMBER KEARNS: I would just like to comment about, as I'm sitting here right now I just thought, should dialysis patients be excluded?

Because, really, that is a patient population that is - and I think of this of all the measures that we're talking about in osteoporosis and it just occurred to me now, because that is a very different patient population.

They have a high risk of fracture. They have a complicated set of problems.

They're probably not really the patient population for the intent of this, but yet they're not specified as an exclusion. So, I might encourage the developers to rethink that, you know.

You could make an arbitrary level of renal function, but you can certainly say patients on dialysis are not the intent of this in general.

DR. SIRIS: If a patient is on dialysis, it might alter your choice of therapy. But at the same time, you need to do the assessment because it may not be a dialysis, I mean, it may not be that it's renal bone disease. It may still be osteoporosis. So, I don't know that I would exclude those patients.

If you had somebody in renal failure who's fracturing and you really don't know what's going on, you would do a bone biopsy. And you might be able to determine that the patient's reason for fracturing was

not the renal failure, but, in fact, was ordinary osteoporosis. And that might influence you to use a drug that you can safely use in a dialysis patient.

I don't know what you would do, but the point is I don't want to make these patients be ignored, because they might have osteoporosis.

MEMBER KEARNS: Well, I agree, but how would this set of tests help you with that?

DR. SIRIS: It would tell you that this was somebody in renal failure, which would force you to think about what was actually going on in that patient.

MEMBER BREEN: And if I could just make a comment when we talk about when we get to potential harm or unintended consequences, I think that's what you're kind of getting into, the unintended consequences if you don't exclude someone who might be inappropriate for bisphosphonate therapy.

But I think we also have to err on the side of some clinical assessment that if we're saying their renal function has to be assessed, the next step is then you use that assessment to drive your clinical treatment of the fracture.

MS. DOMZLSKI: Yes, patients who have had these lab tests within the prior 12 months are excluded from the measure.

So, it's likely the dialysis patient would have had these tests, say, for the 25(0H)D, which you can give them the Vitamin D dose or do the test.

MEMBER KEARNS: Right, but they wouldn't be, I mean, to do a, I mean, I guess you can do a DXA on a dialysis patient. We don't routinely do that.

MEMBER BREEN: But for this particular measure -

MEMBER KEARNS: For this one. But for the constellation of them, you know, I guess thinking of the prior one, the DXA or an

appointment for a DXA in a dialysis patient is a little bit different.

CO-CHAIR GOLDEN: We're getting into specifications. So, we want to - Jessie, do you have anything on evidence?

MEMBER SULLIVAN: I wanted to ask the developers to address why you think that giving Vitamin D while the patient is in the hospital is equivalent to testing their - testing their Vitamin D level.

DR. SIRIS: One of the problems is that the rest of the tests come back the same day. The 25 D level may not come back for a week and a half.

So, when Cathy was busy assessing all of this at the various hospitals, they were saying, gee, you know, waiting for that test to come back is going to be a problem for us. What about if we just start the patient - if we've identified that this is somebody who's got osteoporosis and they're going to have to have follow-up, we'll start them on

Vitamin D and at some future point they can have either the test or somebody can decide that they're adequately treated.

MEMBER SULLIVAN: That's the problem. It's not going to be adequate if you don't -

DR. SIRIS: It was a practical measure to be able to get it done while the patient was still in the hospital.

If you're only going to be in the hospital three or four days and that 25 D test isn't going to come back to the electronic record for a couple of weeks, it becomes a problem in terms of just making it happen.

If you can show that you've started the patient on D, big doses of D, that's a way around dealing with the 25 D level, which is also an expensive test.

CO-CHAIR GOLDEN: So, are you saying you would start therapy without any documentation?

MEMBER BREEN: But again it's based

on the - if I can ask for clarification, that's the dose of D that was settled on. Because again if you ask 12 different endocrinologists their doses of D, this dose was settled on what was considered reasonable in the public health world in terms of recommendations about what certain type of people should take for basic supplementation; is that -

CO-CHAIR GOLDEN: I guess the question is going to be - all right. I guess it is an evidence question, right?

DR. SIRIS: It's very hard to get anybody into trouble starting them on Vitamin D. And I think it would be within the judgment of whoever was taking care of the patient, whether it would be 50,00 units the day they came into the hospital, whether you would put them on 2,000 a day.

I mean, you might take a history and find out that the patient, you know, was on a multivitamin and you might start a lower

dose of Vitamin D, but the point is that you want to assure that you're not missing somebody with D deficiency.

And the way to do that is either to draw the blood, but you can't get it back fast enough, or simply start them.

And by virtue of having them involved in this process, you hopefully will follow up with it.

I think if we require the blood test, if we can require the blood test. But if we do that, the hospitals are going to balk that they can't get the result. It's not feasible. So, again, it's better to start the D or draw the test.

If you've got a fracture liaison service, you'll be able to get the test report, because the fracture coordinator will get it two weeks later.

CO-CHAIR GOLDEN: That gets into solutions, though. That's a different issue.

All right. Any other questions on

evidence? Comments on evidence.

(No response.)

CO-CHAIR GOLDEN: Ready to vote.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: We have high, one. Moderate, 12. Low, six.

MEMBER BREEN: All right. Moving on to the performance gap. Again, for this particular measure, which is looking at laboratory assessment for secondary causes, it doesn't currently exist.

So, we don't really know except in the pilot when they did the pilot, they found that I think it was less than 10 percent of your hospitals were doing these assessments on patients with known fracture in the hospital.

So, I think there's definitely a performance gap that was demonstrated amongst the pilot project.

Does anyone have any other questions or additions?

CO-CHAIR ROSENZWEIG: The pilot, were the hospitals vigorously going back to find out if all of these tests had been done in the past year?

MEMBER BREEN: Well, that was a marker of exclusion, right? So, if they had had the test done in the last year, they weren't included, right?

MS. DOMZLSKI: That's correct. Yes, they did look for that.

CO-CHAIR ROSENZWEIG: No, but in order to exclude them, you need to find out if the tests were done.

I imagine it's kind of hard for hospitals to go back retrospectively to find out if the physicians who had been following the patients had been doing it.

I mean, it sounds like what the thrust of this would be to sort of tell hospitals to do all of these five tests on everybody just to cover their basis.

MS. DOMZLSKI: Well, typically the

patient who's being hospitalized for a fragility fracture is going to have surgery and most of these tests are done as a matter of course anyway. So, they are not out of the ordinary for a surgical patient.

And to address your previous comment about the test group gap, the median level of performance was 9.5 percent.

MEMBER KEARNS: I would just guess that the biggest hangup was the Vitamin D in that.

Seeing all the hip fracture patients at Mayo Clinic, the tests that weren't done until I demanded they be done was the Vitamin D.

They still don't do a calcium or a liver test. The others they do routinely on admission for a fracture. So, I think it's adding a little bit, but not a lot.

Getting back to your question about within the last year, again when we're operating in a hospital facility system and

you had an outpatient one, there is the risk, I guess, of duplication of things, because you don't know what has been done.

I think what the hospitals would do is probably just order that all the tests be done on admission and a Vitamin D pill get prescribed, which is a step in the right direction, but there is, I guess, the risk that these tests were done.

The facility is held responsible. They don't know they were done at your office and not at this hospital.

CO-CHAIR GOLDEN: Are we ready to vote as a group?

MS. BAL: Voting is open.

(Pause.)

MS. BAL: We have high, 11. Moderate, seven. Low, one.

CO-CHAIR ROSENZWEIG: Okay.

MEMBER BREEN: So, importance measure. High priority. So, this stresses whether this is a specific national health

goal priority or data demonstrated a high-impact aspect of healthcare. And, again, this is for this specific measure, which is around laboratory measurement.

I think we've already assessed that missed fracture and prevention of secondary fracture is a high priority.

I think the challenge for this group is to figure out if this measure to do the lab testing is a high priority.

And, again, we come back to you can't initiate the correct therapy unless you know where your patient's basic laboratory testing are. So, I think it's linked, in my opinion, but I'd appreciate other comments on this.

I think we need a cookie break.

CO-CHAIR GOLDEN: We're getting there, yeah.

MEMBER BREEN: As a diabetes specialist, I'm sensing a cookie deficit.

(Laughter.)

CO-CHAIR GOLDEN: Ready to vote? Voting time?

MS. BAL: voting is open.

(Pause.)

MS. BAL: Okay. High, six. Moderate, 11. Low, two.

MEMBER BREEN: Excellent. Okay. Moving on to reliability. Again, the data that we have from this is really from their pilot project that they did.

And we've already discussed that in the last session. So, I think the discussion on reliability in the last session also applies to this measure.

Unless anyone has any comments, we found that it was fairly reliable, I think. Good reliability of their data in their pilot study.

CO-CHAIR GOLDEN: Do you have a comment, Ann? Okay.

Are you ready to vote? Okay. Voting.

MS. BAL: Voting is up.

(Pause.)

MS. BAL: High, six. Moderate, ten. Low, three.

MEMBER BREEN: Okay. Moving on to validity. Again, the data that we have is from the pilot testing. The face validity was assessed by the hospital sites.

The only data element that was more challenging when it came to validity was assessing the laboratory tests in the 12 months prior to fracture.

We've already determined that is a challenge. The rest of the in-house testing seemed to have a good validity score.

CO-CHAIR GOLDEN: Ready to vote?

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. High, three. Moderate, 13. Low, three.

MEMBER BREEN: Moving on to feasibility. This is how feasible it is to do

this. These are inpatients. These are laboratory measures that we check on inpatients.

And so, I think in most of our hospitals we now have electronic laboratory measurements. So, I think that it's a fairly feasible I don't see that many barriers to feasibility in terms of the collection.

I think the challenge comes in the assessment of the 12 months prior. And those challenges, I think, have been discussed.

CO-CHAIR GOLDEN: If no one from the Committee is going to comment, I'm not going to all on you.

So, do we have any comments? You definitely want to say something?

MS. DOMZLSKI: Yes, I just wanted to reinforce what you're saying. This is one thing that has actually been helped by the electronic record.

Particularly in an integrated health system, it's very simple to look back

and find what was done as an outpatient or in the physician's office or at an offsite testing facility in addition to what's done in the hospital. It's been a big improvement.

MEMBER BREEN: I also just want to comment that the abstraction period goes for 30 days post-discharge.

So, even those Vitamin D levels that were drawn and not available at time of discharge will be included in that catch.

MEMBER KEARNS: Doesn't it also just allow for the ordering of the test, not necessarily the performance? Was that part of it, too?

Tests had to be performed or an order for the test at the time of dismissal. So, that could even be an outpatient follow-up of a Vitamin D level ordered as an outpatient.

CO-CHAIR GOLDEN: All right. Vote.

MS. BAL: Voting is up.

(Pause.)

MS. BAL: Okay. High, one.

Moderate, 16. Low, two.

MEMBER BREEN: Usability and use. So, the accountability and transparency of this measure if this is new, which it is, the improvement, are we going to show progress?

Again, we didn't get time to discuss this on our workgroup call since we were so busy thinking about different data elements, but I don't see any unintended consequences of this, but this would be the time to talk about patients not clinically or appropriately excluded from treatment if people worry about a reflexive treatment of patients in order to check a box off.

MEMBER SULLIVAN: I'm sorry, I guess I just wanted to raise again that I think everything we've just said in the last few minutes is a reason to not allow giving Vitamin D instead of ordering the test, because, you know, and I just think it's going to make it less useful, because people will leave the hospital taking a vitamin that they

don't know they really need. So, they won't keep taking it.

And their doctor won't really know they really needed it. So, they won't

MEMBER BREEN: So, you're worried that the potentially substandard treatment of a potentially real Vitamin D deficiency

MEMBER SULLIVAN: Yes.

MEMBER BREEN: could actually add to harm. Because as a

MEMBER SULLIVAN: Because they allow you to get out of ordering the test by giving the vitamin while they're in the hospital.

But, I mean, I think it's still, you know, it's still good, but I just think, you know, you gave people you didn't need to throw them that anyhow.

CO-CHAIR GOLDEN: Is that Patricia on the end? Is your card up? Okay.

MEMBER McCOLLISTER-SLIPP: One question I have as somebody who just finished

a course of Vitamin D, oral Vitamin D, and really liked it and was excited about the energy I got, the doctor told me that I needed to stop because of potential implications for kidney stones, which is a compelling reason.

But is there a possibility that, I mean, and I don't know, I really this is

MEMBER BREEN: I think that the doses recommended here that people get credit for, they were fairly low and benign doses that it would be very hard to invoke some harm at Vitamin D 800 units.

I think when you deal with the bigger doses of Vitamin D, the ergocalciferol 50,000 units in the little gel or green tab that you probably the gel cap, that's where you can potentially have some more clinical negative impacts if you're not treating that person appropriately.

But I think, again, it gets back to, you know, did we give people a way out? Did we make it too easy for them to get their

Vitamin D assessment credit by allowing them to have just the basic, cheap Vitamin D supplementation?

CO-CHAIR GOLDEN: Jessie, can you turn off your mic?

MEMBER KEARNS: Well, I would just like to share a little of experience that I have from trying to do this in my institution before I realized this was a measure.

And I can tell you that from looking at hip fracture patients, that very few patients were actually dismissed on Vitamin D, but a simple education of the orthopedic team and discussion with the endocrinology team we were able to improve that.

Now, we chose a strategy to measure and treat. And that might be because we're the Mayo Clinic and we have access to labs that can be back within 24 to 48 hours, but there was a discussion and promoted by some that every patient who comes in with a

hip fracture should get 50,000 units of Vitamin D on admission.

On dismissal, they're put on a thousand units a day and it's a done deal. You don't have to spend $250 measuring a Vitamin D level.

So, there were alternative strategies at my institution that were discussed. And for a variety of reasons mostly having to do with stakeholders in the orthopedic and endocrine community, we opted for a test-and-treat strategy.

I can tell you that in the hip fracture patient population in our institution depending on how you define Vitamin D deficiency, which again maybe Dr. Siris can help us there, there's still not agreement in all sectors that it's somewhere between 50 to 70 percent even in those people who say they're taking something when they are admitted to the hospital.

So, I think there's a lot of

strategies to achieve the same thing. And I think that, you know, beginning somewhere is a good idea and I don't I'm not averse to just treating people with Vitamin D at these levels.

You won't hurt anyone. Will you optimize some people? Maybe not a couple.

MEMBER BREEN: Most hospitals will I don't want to speak for most hospitals, but I can see panels being developed, right? So, if a fracture panel is developed, Vitamin D will go on it and then you get your assessment, right?

So, let's say they get a reflex of 800 units of Vitamin D and they get the test done and, oh, yes, it comes back two weeks later severely deficient.

One would hope that that will translate along and catch up with the patient down the road if it's part of a panel that's being done.

CO-CHAIR ROSENZWEIG: In Minnesota

in the winter, I imagine there's a pretty high level of D deficiency.

MEMBER BREEN: We should all be taking 50,000 like every day.

(Laughter.)

CO-CHAIR GOLDEN: Are we ready to vote? So, Ann, are you okay with this?

MEMBER KEARNS: Yeah, I would be okay with this. And I was actually one of the people who advocated for just treating at my institution.

CO-CHAIR GOLDEN: Can we vote?

MS. BAL: Voting is ready.

(Pause.)

CO-CHAIR GOLDEN: We have I think someone on the phone is typing. If you can mute, that would be great.

(Laughter.)

MS. BAL: Let's all try again. We only got 16 here.

(Pause.)

MS. BAL: All right. There we go.

So, high, four. Moderate, 14. Low, one.

CO-CHAIR GOLDEN: Okay. Ready for the big picture.

MS. BAL: Voting is ready.

(Pause.)

MS. BAL: We have yes, 16. No, three.

CO-CHAIR GOLDEN: Thank you, Tracy. Now, we go to Bill Curry. And no cookies for you. I'm sorry, Tracy.

MEMBER CURRY: This is Measure 2418, Discharge Instructions, Emergency Department.

So, this is looking at the portion of patients over 50 who have fractures as we've talked about in previous measures, who have been discharged from the emergency room to home, who have received written discharge instructions or their caregivers have received discharge instructions with a need to follow up with a primary care physician, hospital outpatient department or specialists for

possible osteoporosis to reduce the risk of future fracture, or who were contacted by a fracture liaison service.

So, the numerator is patients or caregivers who receive discharge written discharge instructions regarding the need to follow up, or that were seen by, contacted by or linked to a fracture liaison service.

The denominator are patients age 50 or over discharged to home from the ED with one of the ICD-9 codes and soon to be ICD-10 codes for one of the fractures that we've talked about in the Excel spreadsheet that are listed in the SharePoint.

So, the discussion that we had was that we think that certainly this is an important measure. And I think it falls in line with the comments that we've had with the previous two measures that we've looked at.

But the big concern that was raised was that there's great evidence that supports the use of the fracture liaison

service, but there's very little evidence that would support giving the patient discharge instructions to follow up with their primary care physician or other hospital-based or outpatient-based provider for their care after a fracture to get tested for bone mineral density or for treatment.

Initially when we looked at this, we did not have any meta-analyses that helped us with that information, but there were several comments from folks advocating for the review and approval of this study or this measure.

And one of the articles that was cited by those advocates was a study by Ganda in February of 2013 in Osteoporosis International. And this group looked at four models of care for fracture evaluation treatment.

And they went from a Model A which include fracture liaison service, Model B which is similar to fracture liaison service,

but there was an assessment and recommended treatment made to the primary care provider.

The third model was that the patient received education and the PCP received communication through a variety of mechanisms to let them know about the event and their visit to the emergency department and the need for follow-up testing and treatment.

And then the fourth model was some sort of education piece or a recommendation to the patient to be seen by their primary care provider.

And certainly Model A is the best of those models in the meta-analysis that was done with significant improvement in both bone mineral density testing and also treatment.

Model B and Model C also showed some improvement, but there was really no improvement in the providing the patient an education piece and asking them to follow up with their primary care physician.

And so, although we agree that the fracture liaison service as part of this measure is an important part of it, we find no evidence to support on the discharge of a patient just to receive in a discharge instruction that that's going to improve their chances of getting bone mineral density testing or treatment.

So, based on that, it's kind of there's two places that we can go. So, the first part with just the discharge instructions, we think the evidence is low to support that.

With the fracture liaison service contacted at the time of discharge from the ED, we think that the evidence is high to support that.

So, that was our challenge as we looked at this measure for evidence.

CO-CHAIR GOLDEN: So, your committee was mixed or less than happy.

MEMBER CURRY: Well, again, the

evidence is mixed, but we ran out of time. So, we didn't have a lot of time to discuss this.

CO-CHAIR GOLDEN: Okay. Other comments from the Committee.

MEMBER BREEN: I think it would help if maybe the developers gave some insight as to why these two were linked together as almost comparable.

CO-CHAIR GOLDEN: I just want to give the Committee a shot.

MS. DOMZLSKI: Thank you. There is another measure that is currently NQF endorsed. It calls for a transition record to be given to discharge patients with specified elements.

One of those elements is follow-up regarding tests or treatments that need to be done following discharge.

And so, this measure in a large form, addresses the specific wording and information that needs to be in that discharge

instruction that's already endorsed for the other measure.

In addition, there is a publication from 2010 for safe practices and it recommends that discharge systems be in place.

It says, a written discharge plan must be provided to each patient at the time of discharge, it's understandable, and it needs to include, dah, dah, dah, dah, dah, coordination and planning for follow-up appointments that the patient can keep, among other items.

CO-CHAIR GOLDEN: So, just to clarify on the discharge, is that for ER discharge, or hospital discharge?

MS. DOMZLSKI: Just the transfer of the patient care from a hospital to primary care or other community providers. It doesn't specifically state hospital or emergency department, inpatient or ED.

So, that is from the NQF

publication of safe practices. And we feel that this measure in addition to fulfilling those items, rounds out, if you will, and gives the same care to ED patients that the patients in the hospital are going to receive via the other two measures.

CO-CHAIR GOLDEN: I believe, though, we're talking about two different things. It's different to give a patient a discharge instruction with the recommendation to follow up with their primary care physician versus coordinating that care.

When a patient of mine is seen in the emergency department in my institution, before they leave that department they have an appointment and follow-up with me. That's coordination of care.

Or if it's after hours and they can't get that appointment, there's a list that's provided to the medical office assistants in my practice, in all of our practices, about patients that need follow-up.

That's coordination of care.

But as the measure is written, we're not talking about coordination of care. We're talking about giving a handout that suggests that they see their family physician.

I think there's a nuance there and I think it's different.

MEMBER KEARNS: Well, I think if we're just talking about the evidence and the intent of the measure to improve osteoporosis treatment, I think we all agree that anything we do will be an improvement.

But if we're really strictly talking about the evidence that giving information will achieve that, I mean, we have to really look at what's there and it's just not there.

Now, maybe the climate has changed since the original studies were done that are included in the meta-analysis. Maybe it would be more received. But I think for that particular item it's hard to say that there's

good evidence that will change.

And that's very different than the evidence for a fracture liaison service which is outstanding that that works.

And I think the low bar is what people will go for here. And I think that's where the evidence is the weakest.

CO-CHAIR GOLDEN: Tracy.

MEMBER BREEN: I have a question just I also have some guidance from our NQF leadership.

You referenced that it is already an NQF measure that documentation is given about a patient's disease state when they leave. And that your thinking including this, was that you wanted to define and make that language precise as it relates to their fracture in the hospital.

So, are we saying that that measure kind of was already out there in a general thing, we're just making it disease-specific?

MEMBER KEARNS: Right.

CO-CHAIR GOLDEN: This is an ER measure.

MEMBER BREEN: This is an ER visit, right? But isn't that documentation burden still at the level of the ED as well? Meaning if the patient isn't admitted and they visit an ED, they're also required to have some kind of documentation about why they were, you know, why did you come to the ED, you know? What's your follow-up plan?

So, I might be getting off topic, but that's my question.

MEMBER MILLER: I just wanted to clarify there was discussion whether it's ED or inpatient discharge, but the measure is called "discharge instructions, ED."

CO-CHAIR GOLDEN: Do we have other comments on evidence? And I'll ask Janice to put her card down.

All right. Ready to vote.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: We have moderate, seven. Low, ten. Insignificant, two.

CO-CHAIR GOLDEN: Okay. NQF staff.

MS. TIGHE: All right. This measure does not meet the importance criteria and it will not be recommended for endorsement.

CO-CHAIR GOLDEN: So, having said that, are we finished on this measure? Okay. So, we will I think that the measure can be revised and returned and so forth. So, yeah, those are done.

MEMBER HAYDON-GREATTING: So, you're only going to create a measure for people that are going to go home, not to a bridge program, reach program, rehabilitation program post any intervention at the hospital?

Sometimes the elderly patients come in, they have a fracture. The family doesn't feel like they can make sure they get to their rehabilitation center. So, they send

them for three to six weeks, depending on what the doctors have ordered.

MEMBER KEARNS: So, this is from the emergency room.

MEMBER HAYDON-GREATTING: Right.

MEMBER KEARNS: Is that what you're talking about?

MEMBER HAYDON-GREATTING: Right. Yeah, sometimes they're sent from the in that elderly population that kind of needs extra care. So, I wouldn't eliminate it to just a caregiver and home.

CO-CHAIR GOLDEN: You know, the other thing

MEMBER HAYDON-GREATTING: Yes, this measure

CO-CHAIR GOLDEN: And the other thing is

MEMBER HAYDON-GREATTING: Right. I know. But when they go to revise it, I just wanted them to just to think about what's happening out there in the future.

CO-CHAIR GOLDEN: I think that's a good point.

MEMBER HAYDON-GREATTING: Right. Yeah.

CO-CHAIR GOLDEN: I think a large percentage of these people are going to be going to long-term care facilities. Either intermediate or long-term.

MEMBER HAYDON-GREATTING: Right. For short-term since it yeah.

CO-CHAIR GOLDEN: Some sort of liaison that way should be part of the measure.

The other thing just to keep in mind when you revise is that if you're in an ER, that may not be connected to your PCP and you make an internal referral to your liaison service. You get into issues of insurance coverage and approvals and so, that gets real complicated also.

MEMBER BREEN: But I think the bottom line is there just wasn't evidence to

support it. Like you said, you know, I think that until there's some evidence established that's credible

MEMBER CURRY: Or it's written that there is an active attempt to make an appointment with the PCP or orthopedics clinic or another hospital clinic. If the appointment is made

MEMBER BREEN: So, you want more teeth in it.

MEMBER CURRY: Right. Yes. Because the rest of the measure, we thought, stood well.

MS. WATT: Well, we'll be back.

(Laughter.)

CO-CHAIR GOLDEN: It's 5:15 for people on the phone.

MS. TIGHE: Operator, if you could see if anyone on the line has a public comment?

OPERATOR: At this time if you would like to have a comment, please press

star and the number one on your telephone keypad.

(Pause.)

OPERATOR: And there are no comments at this time.

CO-CHAIR GOLDEN: Tell you what. Why doesn't everybody just stand up for a couple seconds? Give yourselves a little break.

I am told we're going to do one more measure before we are given a so, if you want to just walk around or get up for a second, it's

(Whereupon, the above-entitled matter went off the record at 5:20 p.m. and resumed at 5:23 p.m.)

CO-CHAIR ROSENZWEIG: Measure 0056, the diabetes foot exam. The measure steward is the NCQA.

If you'd like to discuss it?

MR. REHM: Sure. Just a quick comment. In contrast to the other five

measures you reviewed, this is only a physician-level measure. So, it's not in HEDIS Health plan.

And, again, it's used in our diabetes recognition program and it's also used in the PQRS program.

It's a fairly straightforward measure looking at foot care and appropriate examination.

CO-CHAIR GOLDEN: So, I can present this. So, the numerator is the denominator are patients of 18 to 75 who by the end of the year had a diagnosis of diabetes. And had an exam during the measurement year or the year prior to the measurement year sorry, had diabetes during that year.

And the numerator is people who this is important. It's a three-part it's a three-part requirement to fulfill the numerator.

You have to have a visual inspection, which I would assume would be a

description of the deformities and so forth, a sensory exam of the monofilament, and a pulse exam during the measurement period. So, there's three things you have to do to pass the numerator.

Now, in terms of the evidence, this is where things get kind of strange. So, I think that the committee or subcommittee discussed this. And I think everybody agrees that some sort of a foot exam or some sort of assessment of risk for the foot is important.

You can talk about different patients at different levels of risk depending on the condition of the foot, but then the issue comes up to do you have to do a sensory exam or the monofilament?

The evidence for the monofilament exam, which is cumbersome and often not done because it's cumbersome, is about is fairly weak.

There are alternative methods. For example, the Ipswich Touch Test which was

in Diabetes Care July 11, is comparable and is a lot simpler to do and to do in the office.

So, we have some concerns about the level of evidence to acquire that a sensory exam be done with the monofilament, per se, which itself would be a potential burden and barrier to completing the exam. So, that would be a question on the evidence on this issue.

The need to, you know, foot exams, obviously diabetes ulcerations and diabetic amputations are a serious problem. Foot injuries are slow and expensive to heal. I don't think anyone would disagree with the evidence for that.

The question of exclusions about this has to be done for everybody, whether or not they have already been shown to have neuropathy or not, is an interesting question. But, again, it's a three-part exam and the question is, is the monofilament an evidence-based requirement that that's the only way to

get this done?

That would be the sum of my comments. And, Sue, do you want to make some comments on that?

MEMBER KIRKMAN: So, just this is going to be a recurring theme, I think, when we talk about the other foot measures tomorrow. And that is that the evidence that exists for ulcer prevention is typically of a very sort of comprehensive program that kind of starts with risk assessment and then, you know, higher risk people get some sort of more comprehensive care. And then, you know, on down the line there are fewer foot ulcers. It's actually mostly foot ulcer prevention.

And one of the problems is trying to isolate out, you know, is there evidence that doing the foot exam with X, Y and Z versus not doing the foot exam, you know, prevents ulcers.

It's just very difficult, because the evidence is all for a more comprehensive

thing. And it's a little bit like the measuring the hemoglobin A1C, except that I think that's, you know, definitely a little bit more clearly linked to the evidence chain.

So, that was just a limitation that we found with all these foot measures is that the evidence for any specific exam is difficult to come by.

CO-CHAIR GOLDEN: And is it a screening test for everybody, or is there a subset?

MEMBER KIRKMAN: Right. Right.

CO-CHAIR ROSENZWEIG: I would just say that I guess I'm a little surprised to hear that the use of monofilament is a difficult or onerous test. It's incredibly easy and much easier than almost any other test that one can devise.

It's certainly a lot easier than using a tuning fork or and it has the advantage of really being able to be a yes or no kind of because the filament bends.

It was developed in the Carville Center for Hansen's Disease down in Louisiana and it has been widely adopted.

It just seems to me it is one way of at least making the sensory examination somewhat objective. Because otherwise, you're either picking the prodding someone too deeply or too little with a needle.

So, and it has been in a number of studies, shown to be a fairly good measure of not evidence for neuropathy, per se, but evidence for clinically significant neuropathy to the foot that might lead to an ulcer.

MEMBER KIRKMAN: Right.

CO-CHAIR ROSENZWEIG: So, as a screening tool.

MEMBER KIRKMAN: Yeah, it's probably a better as good or better predictor of future ulceration than the other tests that are typically done.

CO-CHAIR ROSENZWEIG: Yeah.

MEMBER KIRKMAN: I don't know the

one that you mentioned, but

CO-CHAIR ROSENZWEIG: So, I mean, I don't know why, I mean, it seems to me that if one is going to do a test for clinically to actually screen for clinically significant neuropathy that could lead to an ulcer, that's about the easiest test to do.

And we certainly have lots of patients, you know, we'll get to the gap of care later, but there are large numbers of people who are seen by physicians in their offices with diabetes who never get that never take off their shoes.

CO-CHAIR GOLDEN: As I said, I had looked around at the evidence, looking for the evidence for the test. It's about a level 2B or a level 3 evidence. I didn't find too many.

And I also checked with a couple of my primary care colleagues and other Medicaid medical directors and they were all in agreement with what I just said that it's

cumbersome and difficult and not that useful to them.

So, I don't know. It's just a matter of in the primary care community, you know, the rates haven't been improving. So, are there other ways of getting this done? That's my only comment.

CO-CHAIR ROSENZWEIG: I do it on every patient.

MEMBER SHWIDE-SLAVIN: There's the LEAP screening tool which is very well-defined, very simple to use and is extensively used, I think, within diabetes education programs, diabetes educators, as well as physician's offices using the monofilament.

DR. PACE: So, can I just clarify? Because I presented some evidence. Are you saying that there's evidence that they did not present, or that you you're grading this evidence as low quality? I'm not sure

CO-CHAIR GOLDEN: The evidence, I just I did my own review of the of looking

around. So, looking at the ratings.

As I said, I found an alternative method. That's all I was I was saying it exclusively picks one particular technique. That was my concern.

MEMBER KIRKMAN: And I think, you know, again, like I said, you know, there may not be specific evidence for the foot exam versus no foot exam, but there's evidence for the foot exam identifies people who are at higher risk.

And if you take the people that are at higher risk and you implement, you know, some sort of comprehensive care for them, then there is reduction in ulcers.

I think it's reduction in deep ulcers that is statistically significantly reduced, but, you know, but again it's, you know, sort of like the exam is necessary for that risk assessment. But the exam itself, you know, hasn't really been studied and probably never will be at this point, because

you're not going to randomize people to never take their shoes off versus the foot exam, you know.

CO-CHAIR GOLDEN: Yes.

MEMBER SHWIDE-SLAVIN: I'm just looking at the ADA's 2014 guidelines and there's actually B level evidence on using on the monofilament listed here.

CO-CHAIR GOLDEN: Any other comments?

MEMBER DUDL: Yeah, Bill. This is a question. I think the monofilament is one way to go and it's well-documented.

So, the question is, would this go into the low category because it doesn't cite always to go, or a second way to go and that it forces people in one direction?

I'm just a little unclear about how much this second method degrades the fact that the one does work.

CO-CHAIR GOLDEN: You know, again that's something for the Committee to reflect

on. I think that later on we'll see that there hasn't been great improvement in this area.

So, the question is, is the monofilament going to be a barrier to completing the exams on other alternatives that achieve the intent on doing appropriate doing some sort of assessment of how the foot is performing and so forth?

That's my concern. Are there other ways of fulfilling the intent of what needs to be done?

MEMBER MILLER: I don't think it's a test-specific question or problem. I think it's more an office process problem and a time management problem, the time that it takes for the patient to take off their socks and shoes.

And I know in our practice if we have the patients if we have the medical assistants tell every single patient in the office to take off your socks and shoes, there's a higher rate of completion, period.

And I think some of it is seasonal, too, because this time of year patients say, I'm not taking that off, no, you know, not with my boots.

MEMBER BREEN: A comment about process and there's been a ton of evidence to say that simple process measures make a big change in this, you know, whether you incorporate your office staff to do these things with the LEAP scores, whether if you're on a paper record you put a sticker on the chart with a big foot ahead of time that says, look at the foot.

So, I don't think these should be onerous measures. There's a lot of data out there to say that simple, easy steps can do these.

MEMBER MILLER: Certainly if we keep bombarding the patient every single time, too, that they know when I go in it's the expectation. And I think that we've got to change the expectation for the patient, for

the office staff and for the providers as well.

CO-CHAIR GOLDEN: Again, my comment about the evidence is just on the monofilament piece itself.

CO-CHAIR ROSENZWEIG: The data that was presented in here suggests that, I mean, the monofilament and the biothesiometer vibratory sense probably have equal positive and negative predictive value, but the issue is that the monofilament is so much easier to do. I mean, you don't need complicated equipment.

MEMBER MILLER: And they're a lot lighter weight to carry around in your pocket than the tuning fork.

CO-CHAIR ROSENZWEIG: Yes. I still have one of the original ones, you know, that was produced from Louisiana, you know, that I keep, but now we always use disposable ones, you know, that are available.

But anyway, it seems to me that, I

mean, I think it's, you know, as far as I know it's part of the most guidelines that have been developed.

The ACE guidelines as well, I believe, mention it.

MS. BAL: Voting is up.

(Pause.)

MS. BAL: Okay. We have high, four. Moderate, 13. Low, three.

CO-CHAIR GOLDEN: So, we go next to performance gap. So, performance gap, I think that there is a fair amount of understanding that foot exams are underperformed.

And there are ongoing issues well, obviously there are ongoing issues with diabetic foot care, but that there was no great concern about there not being a performance gap.

MEMBER KIRKMAN: Is this where we talk about the age limit, or is that under validity?

CO-CHAIR GOLDEN: That's under

DR. PACE: It could have come under evidence, what does the evidence say? But you'll talk about that in validity, is the measure specified consistent with the evidence?

MEMBER KIRKMAN: Okay. So, just the evidence is that amputation rates are absolutely the highest in older people. So, and very devastating, very costly to Medicare, et cetera, et cetera, et cetera.

DR. PACE: So, let's hold that for validity. Let's talk about this performance gap. Any additional

CO-CHAIR ROSENZWEIG: Any comments?

(Pause.)

CO-CHAIR ROSENZWEIG: Then let's vote.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. The results are high, 14. Moderate, five. Low, one.

CO-CHAIR ROSENZWEIG: So, we'll

move on to importance of the measure.

Bill.

CO-CHAIR GOLDEN: It's a major issue, major problem in diabetes. Prevention of foot ulcers would be a nice thing.

CO-CHAIR ROSENZWEIG: Let's vote.

MS. BAL: Open.

(Pause.)

MS. BAL: Okay. We have high, 17. Moderate, three.

CO-CHAIR GOLDEN: All right. Reliability. This one gets a little more tricky. I'd like to hear a little bit from the developer as well in the sense of, you know, again it requires three things to happen.

And I guess the question is, are the data extracted consistently? And is the documentation consistent? And would you have, quote, a normal exam be considered adequate? And are there specific things that have to be documented to pass the exam, to pass the

measure and the numerator?

And it's unclear how this gets extracted to pass the measure.

MR. REHM: Can I respond? So, again, there's the measure and then we have a program.

And the way our program works is that clinicians get their sample, they look at the patients, they look in their medical record, they can either extract from their EHR, registry, whatever they wish to use, and they would be looking that those three things occurred.

Not or, not this and or that and or this, but just do these three things and you've done a foot exam. And that's the measure.

CO-CHAIR ROSENZWEIG: No, I just think that, yes, it is very important to avoid a scenario where a physician has a box that says "foot exam" and they check it off, which is

MR. REHM: In our program, that's not the way our program works.

CO-CHAIR ROSENZWEIG: Yes.

MR. REHM: I'm not speaking about other programs that may have that dimension to it.

CO-CHAIR ROSENZWEIG: Well, that's why I'm saying

MR. REHM: That's their choice.

CO-CHAIR GOLDEN: One concern was it has not been tested in a primary care community, only in folks who want the recognition.

I don't know if anyone from a plan who collects this data they're in the primary care recognition program as opposed to a general population.

So, I was just curious if health plan has done reviews of this measure, I'm curious how the extractions have gone.

MR. REHM: It's not a health plan measure. I doubt they would want to collect

it just independently, because it's not in our domain.

People are free to ask health plans. I'm just saying it's not a measure that's used in that setting.

CO-CHAIR GOLDEN: So, it's used in is it used in PQRS?

MR. REHM: Yes.

CO-CHAIR GOLDEN: Have there been any data validity with PQRS?

MR. REHM: The PQRS data is included in the submission. It looks very much like the same kind of data that we see, because it's a self-selected group of physicians deciding to report from a constellation of measures, measures appropriate for their practice. Generally speaking, it would be people who take care of patients with diabetes.

CO-CHAIR ROSENZWEIG: Just wanted to ask does the measure specify which pulse is

MR. REHM: No, it leaves that open. So, in the medical record it could probably read a short note, you know, looked at the foot, this is what I found, did a monofilament, you know, here's the result and took a pulse.

CO-CHAIR ROSENZWEIG: Because the way I read it, it suggested that perhaps you could measure the question was does it have to be the two it wasn't clear that it had to be the foot pulses from the way it was defined in the beginning, but I assume that was the

MR. REHM: Yes, no. If the

CO-CHAIR ROSENZWEIG: In other words, you couldn't do a femoral pulse and get credit for this or

MR. REHM: No, it's a foot pulse.

CO-CHAIR ROSENZWEIG: Okay. I mean, it sounds silly, but

MR. REHM: Yes.

CO-CHAIR ROSENZWEIG: Okay. So, should we vote on reliability then, I guess?

MS. BAL: All right. Voting is open.

(Pause.)

MS. BAL: Okay. The results are high, three. Moderate, 13. Low, four.

CO-CHAIR ROSENZWEIG: Validity.

CO-CHAIR GOLDEN: So, the question here would be are the specifications consistent with the evidence? Is there sufficient specificity in the codes? And is the age inclusion consistent with the evidence?

That gets to your age question. That's part of the validity question.

CO-CHAIR ROSENZWEIG: Janice.

MEMBER KIRKMAN: Is this where I can talk about age?

CO-CHAIR ROSENZWEIG: Sue, go ahead.

MEMBER KIRKMAN: Sorry. This is where I can talk about age? Yes, I think the upper age limit is a big problem. I don't see

any justification for it.

I can see, like, microalbumin screening where you're talking about a complication, you know, 10, 15, 20 years down the line in a 90-year-old might not be worthwhile, but, you know, foot ulcers can develop relatively quickly, have a huge impact on quality of life and mortality and costs and so forth.

So, I don't know what the history was behind this other than a lot of the measures seem to be 18 to 75, but I don't think the upper age limit is justified.

I think it's actually kind of almost discriminatory. I mean, it's the people that need it the most that will be excluded.

CO-CHAIR ROSENZWEIG: I think that's a good point. I mean, decubitus ulcers occur in greater amounts in most elderly patients.

MR. REHM: Can I respond to that?

CO-CHAIR ROSENZWEIG: Yes.

MR. REHM: Sue, thanks for that. You know, this is an interesting moment where you have an artifact of a program, because remember those were created around a program.

And just not that this is a health plan-level measure, but from the health plan side we had so many different indicators. We have 10 indicators for health plan measurement and diabetes and some of them you don't want to be doing over 75.

And we just looked at all of them and tried to get at essentially the best common denominator on age. We don't include foot exam in that. So, I want to make sure that's separate.

So, this is a classic case where we're comfortable having the measure endorsed with no upper age limit. In terms of the use and the program that we happen to have, we would probably constrain it because we're looking at A1C is less than A, A1C is greater,

you know. That's our choice to stratify, you know.

We created the measure. We have the IP on the measure. NQF endorses the measure with an upper age limit of none, right? That's fine. We can use the measure in our program accordingly, as does any measure user out there.

So, you can measure something and not you can choose to measure different components of that population or stratify it to meet the needs of your thing.

People don't necessarily

DR. PACE: I need to weigh in on that from an NQF standpoint. NQF endorses a measure as specified, and that's what's the NQF-endorsed measure.

MR. REHM: That's fine, yes.

DR. PACE: We don't have control over how people implement it, but generally if it's not implemented how it's endorsed, it wouldn't be implementing the NQF-endorsed

measure.

So, but again, you know, NQF only can control what it endorses and what the specifications are.

CO-CHAIR ROSENZWEIG: But endorsing a measure with an upper limit doesn't necessarily mean that we're saying that you shouldn't do it at higher age levels.

DR. PACE: Well, you have to think about, you know, that's why we have the evidence criterion, that's why we look at specifications and validity is that this is supposed to be an indicator of quality of care.

And so, if you all are saying the evidence indicates that this process should be performed on patients regardless of age, then it wouldn't be logical to then endorse a measure that you thought didn't match the evidence.

So, we would, you know, suggest that, you know, if this is how the evidence

falls out that there should be no upper limit, then you can you have some options.

You can ask the developer if they're willing to, you know, change the specification. I mean, that's a very limited thing, but, you know, and generally it doesn't happen during an endorsement process, or you can, you know, vote up or down on the measure as it's currently specified.

But I think you need to have more discussion in terms of whether you agree on this should not have an upper limit and then

MEMBER SULLIVAN: So, if I understand I just want to clarify we could ask the developer, please change the limit. And the developer could say, okay. And we could proceed with the vote even though that's not usually what happens, because it's self-specific. You'd let us do that?

DR. PACE: Yes, but it would be up to the developer to say whether they could do

that at this point in time, because a lot of times they have implications and have to go back to their committees and their constituency.

MEMBERS SULLIVAN: No pressure.

MR. REHM: Well, I think you'll recall I made the recommendation.

MEMBER BREEN: If I can just weigh in, I think this is a really interesting opportunity.

Because if you look at the other measures, the reason we have the age limit is for patient safety, right? We put those because we don't want to hurt old people, right?

And the irony here is by having an age limit, we may end up hurting old people, because we're basically implicitly stating that they're out of the view box as it were. So

MEMBER KIRKMAN: Yes, we don't either we don't want to hurt old people or

there may not be benefit once you reach a certain, you know, limited life expectancy.

But here, I think there is no harm and there potentially is benefit, you know, unless you're going to die tomorrow.

So, I mean, yes, I mean, I would hope you would be willing to, because I wouldn't want to vote down the measure based on this.

MEMBER BREEN: Especially with the aging population. When you look at the map of those numbers, just the total N of patients, we're going to be over 75 in the next few years.

MEMBER KIRKMAN: So, more than half the people with diabetes are over 65. I'm not sure about over 75, but it's a big chunk.

CO-CHAIR ROSENZWEIG: It's a big number. All right. Does that mean we have to actually create an amendment or something?

DR. PACE: So, we've heard from Bob that NCQA is willing to remove the age limit.

Is there any objection from the steering committee?

(No response.)

DR. PACE: Okay. Then why don't you proceed with the rest of your voting

CO-CHAIR GOLDEN: I have a secondary question.

DR. PACE: Okay.

CO-CHAIR GOLDEN: The other, I mean, that's one issue. The other issue about validity, again this is I have a question about the exclusions in that somebody with already known neuropathy or foot issues or already under care, would they be excluded?

Or would that be if they were I guess if they're seeing podiatry, they would be in the numerator automatically?

DR. REHM: The specification was designed around ambulatory care and people going though that. So, that is not an exclusion currently. I mean, I don't think we presented an exclusion for that.

CO-CHAIR GOLDEN: Because if somebody already has known already has a known abnormality, to continue to repeat the testing

MR. REHM: Well, they may have an abnormality on one limb, not the other. I mean, I don't know if there's I think you get into sometimes we say do we specifically put an exclusion for a double amputee?

And some people say, well, actually you still need to, you know. It's kind of where do you start and where do you stop.

CO-CHAIR ROSENZWEIG: I think under the circumstances usually if the person is totally anesthetic in both feet, then you start testing further up on the leg basically and document the level.

I mean, that's what a lot of people would normally do.

MEMBER KIRKMAN: I think there is zero evidence for that, though. I mean, I

agree. I mean, there are if they can't feel that monofilament at all, they're already so high risk that, you know, I'm not sure it matters whether they start feeling it at their knee or halfway to their knee.

MEMBER McCOLLISTER-SLIPP: Again, just chiming in on the patient perspective as somebody I have neuropathy. I have pretty good sensation in my feet, but I get pain.

I mean, the level of sensitivity does fluctuate from visit to visit and it, I mean, it can often fluctuate with significant episodes of, you know, high glucose around really stressful events or something.

So, I do think there would be merit in repeating it maybe not every time you see the doctor, but once a year or something.

So, again, this is just anecdotal. It's not based on the evidence presented, but I don't think it's unreasonable given how inexpensive this particular test is.

CO-CHAIR ROSENZWEIG: Any other

comments?

(No response.)

CO-CHAIR ROSENZWEIG: Okay. Let's vote on validity then.

CO-CHAIR GOLDEN: And, again, voting on the validity with the understanding that there will be an amendment, correct?

CO-CHAIR ROSENZWEIG: Correct. Correct. Yes.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: So, high, eight. Moderate, nine. Low, two.

CO-CHAIR ROSENZWEIG: All right. So, well go on to feasibility.

CO-CHAIR GOLDEN: Again, this is extent to which the specifications include measure logic, data readily available, could be captured without undue burden and implemented for performance measurement.

We've kind of gone around in circles about that already.

CO-CHAIR ROSENZWEIG: Yes, Bill.

MEMBER CURRY: So, in the PQRS measure it's just a foot exam. Neurologic examination of the foot and ankle. And yet, this measure has three parts to the examination.

And if a provider or if practices are going to try to capture this information from their EMR, they'll have to have some way to create or accommodate an element for each of those three parts of the measure, or they're going to have to do chart reviews.

So, I just I think that's a problem in terms of the feasibility especially for smaller practices that perhaps don't have the resources to do this kind of work.

Even for larger practices it's going to be a chart review, because most of our EMRs don't have am accommodate and element set with those three pieces in it.

MEMBER BREEN: If I can just comment, the feasibility sounds a lot like the

ophtha report feasibility, right? So, I mean, this is the exact same discussion we had about how to pull those ophtha reports, the ophthalmology reports of the diabetic eye exam, right?

So, anyone who has gone through NCQA certification in their practice knows there are two roadblocks. So, again, documentation and the eye exam.

So, I think the same discussions we've had about that topic play right in here, because it's you're right. There are very few EMRs that have those discrete data fields that you can pull that data from.

CO-CHAIR ROSENZWEIG: I could guarantee, though, that if a measure like this is approved, then the EMRs would include that granularity very quickly.

MEMBER CURRY: And I do believe that this is a better overall assessment of the patient's lower extremity than what the PQRS is going to measure.

So, just a lot of difficulty collecting it until the EMRs catch up.

CO-CHAIR ROSENZWEIG: Yes, Janice.

MEMBER MILLER: Okay.

MEMBER McCOLLISTER-SLIPP: Yes, and my question was primarily around data extraction as well.

I mean, I just see and maybe I'm looking at the wrong list of codes, but I only see ICD-9 codes in what must be more like CPT codes or something since it's a procedure.

I mean, I would think that this would be relatively cumbersome to extract. And as somebody who works with EHR companies, I think they might take a little bit longer to come up with some sort of composite measure that would be built into the base.

MEMBER KIRKMAN: Yes, it's part of the physical exam. So, it would just be part of the E&M visit. It wouldn't be a separate CPT code.

MEMBER MILLER: I was going to make

the same comments about it being the same as having a distinct field for retinopathy screening, but I also think exactly what you said, Jamie. This is the only thing that's going to drive EMR developers to create a distinct field.

And I think if we look at what is our overall goal of this, our overall goal is to drive the quality improvement and to have this conducted and recorded so that it can be measured.

So, rather than saying, well, give the rubber stamp that it's too difficult to record, I think that we need to push the envelope on this.

CO-CHAIR ROSENZWEIG: Just as an aside, you have no idea how far behind the developers are in meeting these opportunities.

Any other comments? Oh, yes, Patricia.

MEMBER McDERMOTT: That's what I was going to say about abstraction and

thinking that an EMR modification is going to happen quickly.

So, you're developing a measure that's going to require manual chart review for quite some time. And I thought I heard that there is another measure, PQRS, that's going after the same concept. It's just not as granular. So, I guess at some point we talk about harmonization, yes.

And I don't know whether that other measure is already endorsed by NQF, but that's been one of the things that has driven a lot of these discussions as well is things that are basically going after the concept, same concept and how does a provider then figure out which thing to do. Just a thought.

CO-CHAIR ROSENZWEIG: Harmonization, I think, is tomorrow.

MS. TIGHE: We actually so, we're moving the two APMA measures to the call that we have scheduled for March 12th. The developer had to leave. And then we'll

discuss the 0519, the CMS measure tomorrow.

0416 and 0417, the developer had to leave. And so, he has asked that we discuss these measures on the call that we have scheduled for March 12th. It's from 1:00 to 3:00 Eastern. I believe you all have calendar appointments already.

We had hoped to give it back to you, but unfortunately we won't be, but 0519 we'll discuss tomorrow morning.

CO-CHAIR ROSENZWEIG: Any other comments on feasibility?

MR. REHM: Just a quick one. There may be more than one foot care measure in PQRS. This measure is in PQRS.

I believe last year they put an "or" instead of an "and." We didn't catch it. It's an "and" in future world it's been approved. So, I just wanted to let you know that there is concordance with that.

Now, remember that program uses either G codes of CPT 2 codes to do that. The

infamous check the box, that's their choice for how they do that measure.

So, to the extent that you believe that physicians would be faithful about doing the exam and then doing that and then that's the method of getting that data, that certainly is more feasible.

People have issues with the kind of the integrity underlying it. So, that's not discussion we want to weigh in on, but okay.

Bill.

MEMBER TAYLOR: It is a discussion that we ought to have though, right? I mean, if there are if there's no good evidence that actually putting a measure like this in place actually results in an outcome that we're trying to achieve and if this is pushing the envelope in terms of what developers would have to do on EMRs and so on, and if this questions even if you can't raise them, that we could about is this actually going to

result in physicians really doing this work or merely checking some box or doing something else rather than doing something that's going to result in the outcome we're looking for, and if there's the opportunity cost if you do this, you don't do something else, and if there's pushback in the physician community that we're requiring them to do things where there isn't evidence supporting it, well, then certainly this is exactly the kind of thing that we should not support and go ahead on.

MEMBER MILLER: This is also something that doesn't need to be done by a physician or nurse practitioner. You know, we have I've trained nurses to do this and to document it in notes.

CO-CHAIR ROSENZWEIG: I think there is evidence. I would disagree about the issue of evidence.

I think as we've mentioned in the ADA guidelines, they're talking about Level 2 evidence; is that correct B evidence, yes.

Yes.

So, I'm not sure the issue, obviously, we're talking about feasibility. So, the issue here is whether or not data capture will be feasible.

We have it in our electronic medical records. I just don't see why it would be a difficult thing to capture this kind of information, myself.

DR. PACE: So, we're on feasibility. So, are we switching back? Does someone want to go back to evidence or

CO-CHAIR ROSENZWEIG: No, no, no.

DR. PACE: Okay.

CO-CHAIR ROSENZWEIG: But someone mentioned Bill mentioned the fact that there wasn't evidence for it. So, I just

DR. PACE: Okay. Thanks.

MEMBER MILLER: And if we think about the process of it going back to the process again even that we don't have distinct data fields for it, practices have figured out

how to do it for PQRS for financial incentives.

So, if they're figuring out a process to do it without a distinct field, they're figuring out a process.

MEMBER SULLIVAN: I wanted to ask Bob if you could just clarify. I got confused by the last thing you said.

So, not the other measure that's in PQRS, but this measure, your measure is in PQRS. So, it is specified with G codes, but we don't have them?

MR. REHM: It's the no, it's included. We don't we don't give you the codes for that program. That's the program choice to use those codes, I guess.

And because we have limited data from the PQRS program, we presented our data from our own recognition program. So, there are CPT 2 codes that are, in this case, that capture this requirement.

I can look them up on our

specifications. I mean, they're there.

MEMBER SULLIVAN: I thought we were being asked to endorse a measure that was being used in two places in PQRS and then but we don't actually have the PQRS specifications; is that

MR. REHM: So, I'm sorry Helen is not here. And maybe some folks from NQF can speak to the issues around endorsement around these coding, check-the-box approaches, because and so, that's why we specify the the measure intent is to go after these three things. And how you can collect that in different programs is up to the program developer.

And I'm so, I just will leave it there and the Karens can maybe respond.

DR. PACE: So, you're bringing this measure to us with the medical record specifications, not the G code specifications.

MS. JOHNSON: So, just FYI, the spreadsheet that you guys submitted with your

measure has a G code for the foot exam. So, I think you did provide it to us.

MR. REHM: Yes, we added that in as a concession to those clinicians who are using it, but that's not in our program, that's not this dominant collection. It was a courtesy, if you will, to help those that were 40,000 physicians in the PQRS program, 3,000 or 4,000 in our particular program.

CO-CHAIR ROSENZWEIG: Well, if it's in PQRS, it presumably has already been approved by the NQF.

DR. PACE: Not all measures in PQRS have been approved by NQF. And this one was the difficulty is that some of the measures that were originally endorsed did not have the testing. And I guess the testing data you've been presented with is from the recognition program using the specifications for the medical record abstraction.

So, I think what you need to do is think about the measure as Bob has described

it, the medical record abstraction. That's what the testing is from.

And I will have to clarify if there's any implications for the PQRS program or how to deal with that. We can come back to that tomorrow with Helen.

MEMBER SULLIVAN: I think there will be implications for our discussion of the other measures tomorrow.

I wonder, Karen, is it possible you could show us where are these here?

Do discuss them.

CO-CHAIR ROSENZWEIG: Well, these are the ones that have been tabled? Okay.

CO-CHAIR GOLDEN: The ones that were tabled were from the podiatrists. The PQRS measure is tomorrow.

CO-CHAIR ROSENZWEIG: Oh, okay.

MEMBER KIRKMAN: The podiatry measures are PQRS measures.

(Pause.)

CO-CHAIR ROSENZWEIG: This is the

NCQA, okay.

So, I don't know what the specification of that particular code is. Does that include the three parts of the

MEMBER CURRY: It says, foot exam performed includes examination through visual inspection, sensory exam with monofilament and pulse exam. Repot when all of the three components are completed.

CO-CHAIR ROSENZWEIG: So, it's consistent, yes. Okay. All right. Okay. So, I think let's vote on feasibility at this point.

(Laughter.)

MEMBER SULLIVAN: Including the HCPC code specification for PQRS, because that's how it was given to us, right? Okay.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. We have high, one. Moderate, 15. Low, three.

CO-CHAIR ROSENZWEIG: So, we go to

usability. And usability, again, is the potential for potential audiences could use or will use performance results for accountability and improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations. So, the impact in value for quality improvement.

CO-CHAIR GOLDEN: So, the usability, that's in red from what the Workgroup determined?

CO-CHAIR ROSENZWEIG: Yes, I think that it was there was, I mean, a few issues here and there, but generally the sense was that it was a usable measure.

CO-CHAIR GOLDEN: Let's vote oh, wait. No, someone has their Jessie.

(Laughter.)

CO-CHAIR GOLDEN: Nobody wants to slow us up now. All right. So, let's vote on usability.

MS. BAL: Voting is up.

(Pause.)

MS. BAL: Okay. So, we have high, seven. Moderate, nine. Low, two.

CO-CHAIR ROSENZWEIG: All right. So, we're now voting on the overall measure with the caveat no upper age limit, and then we're presumably also having this potential for certainly for NCQA, but that the alternative PQRS could be

DR. PACE: Well, I think, you know, given that the specifications were provided and the comment to vote on it with those specifications, but we'll just clarify it tomorrow.

CO-CHAIR ROSENZWEIG: Okay. Thank you.

MS. BAL: Voting is open.

(Pause.)

MS. BAL: Okay. We have yes, 16. No, three.

CO-CHAIR ROSENZWEIG: Okay.

MS. TIGHE: All right. So, a little bit of housekeeping. We will plan to

start at 8:00 a.m. tomorrow so that we don't run into this situation with you all running to the airport to catch your flights.

So, thank you all for soldiering through today. I know it's been a really long day. We really appreciate it to our developer colleagues. Also, we very much appreciate you sticking around for this. The two audience members remaining, also, thank you.

We'll have breakfast at 7:30 for all of you. Please enjoy your evening. I'm sorry we kept you so long, and we look forward to talking to you all again tomorrow.

MS. BAL: And please leave your vote clickers next to your name tags. I'll come get them. Thank you.

MS. TIGHE: Feel free to leave papers in the room or anything that you want to revisit tomorrow.

(Whereupon, the above-entitled matter went off the record at 6:14 p.m.)

**F**

## Q

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Endocrine Measure Endorsement

Before: NQF

Date: 02-26-14

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