THE NATIONAL QUALITY FORUM

+ + + + +

STEERING COMMITTEE ON
NATIONAL VOLUNTARY CONSENSUS STANDARDS
FOR PATIENT OUTCOMES

+ + + + +

MEETING

+ + + + +

TUESDAY
APRIL 20, 2010

+ + + + +

The Patient Outcomes Steering Committee met in Salon 1 in the Marriott Bethesda Hotel, 5151 Pooks Hill Road, Bethesda, Maryland, at 10:00 a.m., Joyce Dubow and Lee Fleisher, Co-Chairs, presiding.

MEMBERS PRESENT:

JOYCE DUBOW, MUP, Co-Chair
LEE FLEISHER, MD, Co-Chair
RUBEN AMARASINGHAM, MD, MBA, Member
LAWRENCE M. BECKER, Member
E. PATCHEN DELLINGER, MD, Member
ANNE DEUTSCH, PhD, RN, Member
BRIAN FILLIPO, MD, MMM, FACP, Member

LINDA GERBIG, RN, MSPH, Member
EDWARD F. GIBBONS, MD, Member
LINDA GROAH, RN, MSN, CNOR, FAAN, Member
PATRICIA K. HAUGEN, member
DAVID HERMAN, MD, Member
DAVID S.P. HOPKINS, MS, PhD, Member
DIANNE V. JEWELL, PT, DPT, PhD, CCS, Member

DAVID A. JOHNSON, MD, FACP, FACG, FASGE, Member
IVER JUSTER, MD, Member

MEMBERS PRESENT (Cont'd):

BURKE KEALEY, MD, FHM, Member

PAULINE McNULTY, PhD, Member

LEE NEWCOMER, MD, MHA, Member

VANITA K. PINDOLIA, PharmD, BCPS, Member

AMY K. ROSEN, PhD, Member

BARBARA J. TURNER, MD, MSED, MA, FACP, Member

BARBARA YAWN, MD, Member

ALSO PRESENT:

HEIDI BOSSLEY, MSN, MBA, SENIOR DIRECTOR,
PERFORMANCE MEASURES

HELEN BURSTIN, STAFF

HAWA CAMARA, STAFF

SARAH FANTA, STAFF

SEAN O'BRIEN, MD, CONSULTING STATISTICAL
REVIEWER

REVA WINKLER, MD, MPH, PROGRAM CONSULTANT

SEAN O'BRIEN, MD, Consulting Statistical Reviewer

NQF STAFF:

HEIDI BOSSLEY

HELEN BURSTIN

HAWA CAMARA

SARAH FANTA

REVA WINKLER

T-A-B-L-E O-F C-O-N-T-E-N-T-S

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

10:09 a.m.

CO-CHAIR DUBOW: Good morning, everybody. I am Joyce Dubow, Co-Chair.

CO-CHAIR FLEISHER: I am Lee Fleisher, the other Co-Chair.

CO-CHAIR DUBOW: We are a little bit late, and we will go around the room, introduce ourselves, declare whether we have any conflicts, and then we will review the agenda for the next two days.

We would very much appreciate it if the name tags could be directed toward us so that everybody gets called by the proper name, and also, please, when you introduce yourself, if you were Chair of one of the Technical Panels, please also let us know that so that we will know who is who.

I am Joyce Dubow from AARP.

CO-CHAIR FLEISHER: I am Lee Fleisher from the University of Pennsylvania, Chair of Anesthesia.

DR. WINKLER: I am Reva Winkler. I am a Project Consultant to NQF.

MEMBER AMARASINGHAM: I am Ruben Amarasingham. I am a physician at Parkland Health and Hospital System.

MEMBER JOHNSON: David Johnson, a gastroenterologist from American College of Gastroenterology, and I was the GI TAP Chair.

MEMBER JUSTER: Iver Juster from Outcomes and Informatics at ActiveHealth Management. No disclosures.

MEMBER AMARASINGHAM: No disclosure.

MEMBER JOHNSON: David Johnson, no disclosure.

MEMBER KEALEY: I am Burke Kealey. I am a hospitalist for HealthPartners Integrated Delivery System in Minneapolis, and also an officer of the Society of Hospital Medicine. No disclosure.

MEMBER McNULTY: Hi. I am Pauline McNulty from Johnson & Johnson. No

disclosures.

MEMBER GERBIG: Hi. I am Linda Gerbig from the Association of Perioperative Registered Nurses. No disclosures.

MEMBER BECKER: Hi. I am Larry Becker. I am the Director at Xerox Corporation, and I am also on the Board at The NQF.

MEMBER TURNER: Good morning. I am Barbara Turner, American College of Physicians. No disclosures.

DR O'BRIEN: Hi. I am Sean O'Brien. I am a statistician at Duke Clinical Research Institute, and I am here as a consultant for NQF as a fiscal reviewer. DCI is involved with measures to do with STS. So when those are discussed, I will not be here as a consultant for NQF.

MEMBER FILLIPO: Hi. I am Brian Fillipo, the Vice President for Medical Affairs at Bon Secoeur, St. Mary's, and I have no disclosures.

MEMBER YAWN: Barbara Yawn, family physician, health services researcher, and I was Chair of the TAP, Pulmonary TAP, and I have no disclosures.

MEMBER HAUGEN: Pat Haugen, consumer, National Breast Cancer Coalition. No disclosures.

MEMBER PINDOLIA: Vanita Pindolia, pharmacist, Henry Ford Health System in Detroit, Michigan, medication management programs, and I have no disclosures.

MEMBER GROAH: Linda Groah, CEO of APRN, and I have no disclosures.

MEMBER JEWELL: Dianne Jewell. I am a physical therapist representing the American Physical Therapy Association. I was the Chair of the Bone and Joint TAP, and I have no disclosures.

MEMBER DEUTSCH: Anne Deutsch. I am a clinical research scientist, Rehab Institute of Chicago. No disclosures.

MEMBER NEWCOMER: Lee Newcomer

with United Health Group, and I run the cancer and cells divisions there. I chaired the Cancer TAP, also the former Chairman, Park-Nicollet Health Services in Minneapolis.

MEMBER HOPKINS: David Hopkins, Director of Quality Measurement, Pacific Business Group on health. I have no conflicts to disclose.

MEMBER DELLINGER: Patch Dellinger, Professor of Surgery and Chief of General Surgery at the University of Washington. I chaired the IV TAP. I don't think I have any disclosures relevant to here. I have consulted for almost every pharmaceutical firm that makes any antibiotics over the years, but I don't think that is relevant to what we are doing here.

MEMBER ROSEN: Amy Rosen. I am health services researcher at Boston University, School of Public Health and School of Medicine, and also at the VA.

MS. CAMARA: Hawa Camara, analyst

at NQF.

DR. BURSTIN: Good morning, everybody. I am Helen Burstin, Senior Vice President at NQF.

MS. BOSSLEY: good morning. I am Heidi Bossley, Senior Director for Quality Measures at NQF.

CO-CHAIR DUBOW: Are there any Committee members on the phone? Is Shelley? Okay.

(Off-mike introductions.)

MS. FANTA: Hi. Sarah Fanta, research analyst with NQF.

CO-CHAIR DUBOW: Are there any members of the public on the phone who would like to introduce themselves, please? We heard a few of you. Rita?

DR. GALLAGHER: Hi. It is Rita Munley Gallagher with the American Nurses Association.

CO-CHAIR FLEISHER: Bruce, are you there?

CO-CHAIR DUBOW: I thought we heard somebody else this morning.

DR. PATTON: Mary Patton from the Association of American Medical Colleges.

DR. HALL: Bruce Hall from the American College of Surgeons.

CO-CHAIR DUBOW: Okay, thank you very much.

DR. WINKLER: Welcome, everyone. The last six months since we met in October has been a very busy and intense time. As you all participated in, we had two conference calls in March to evaluate an initial set of 12 measures, and I am going to give you the results of all of that, but at the same time, numerous conference calls and meetings were held with the Technical Advisory Panels preparing for this meeting.

So thanks to everybody for your participation. But let's go on to what happened in the past, so we know where we are.

Out of the 12 measures that you

evaluated over the conference calls in March and 17th and 24th, you recommended eight of them to go forward for endorsement.

Six of the measures were recommended straight out for regular endorsement, and they are the Intensive Care In-Hospital Mortality Rate, as well as the Intensive Care Length-of-Stay measure paired with the ICU.

The complication rate for ICD and the 30-day readmission for PCI, and then the composite measure for AMI discharge care transition, and also the heart failure discharge care transition composite measure.

So a majority of the Committee members recommended that these go forward for endorsement.

Any questions from anybody on the Panel? Okay, now you did recommend that two of the measures go forward for time limited endorsement, and these are the two measures around pulmonary rehabilitation. The first is

the health related quality of life, and the second is functional capacity. Any questions about those?

We have had conversations with the developers of these two measures, and they are in agreement with doing the testing of these measures within the 12-24 month time frame that NQF requires. So that is where we are.

MEMBER YAWN: Is it 12 or is it 24?

DR. BURSTIN: Twelve months is our new policy. This project hit it smack dab in that transition point. So I think we will allow them to go up to 24 months, but preferable, the sooner the better, and they know that.

MEMBER YAWN: That is good. I was just going to make a complete 24 months.

DR. WINKLER: So any other questions on that? Okay.

The four measures that were not recommended -- but I will tell you that those

were close-ish, but these were the individual measures that are part of composites. The majority of -- About half the Committee really only recommended these measures for the composite only, and then a goodly -- and then several more not at all.

So as stand-alone measures, these four were not recommended, but they are part of the composite, which you did recommend. Questions on that?

Okay, so that is where we are. Given that --

MEMBER JUSTER: So they are still going to be reported?

DR. WINKLER: These measures are part of the composite. Correct. Just but you did not recommend them as stand-alone, independent measures.

Given that this is sort of a decision endpoint, we would like the opportunity for public comment on them.

CO-CHAIR DUBOW: We will have

public comment in just a minute. I just want to remind you what happens to our recommendations. Do you want to explain that?

DR. WINKLER: Sure. What is going to happen next is we have drafted a report that describes the discussions and the evaluation of these 12 measures. We are finalizing it to be released for public comment in May. All right?

We expect to get feedback from NQF members as well as members of the public at large. Those comments will be collated and organized by measure.

We will also be asking comments on the measures not recommended, and then we will be coming back to you sometime in June, probably around the third week of June, to look at those comments to see how -- This is feedback for your decisions, to see if it changes your mind, gives you a different way of thinking about things, brought up issues you hadn't considered, whatever.

This is sort of an opportunity to get feedback on the decisions you made, and at that time make any revisions which you feel might be necessary based on those comments.

After that, the revised draft report will be sent to NQF members for voting. This is planned for the month of July, and then the results of voting will go to the Consensus Standards Approval Committee in August.

The Consensus Standards Approval Committee of CSAC is a subcommittee of the Board of Directors whose charge is to look at the process of evaluating and recommending these measures as well as the measures themselves on behalf of the Board. Then the Board endorsement is scheduled for September.

So that is where we are going to go with these measures going forward. So your role is not finished with them totally. We will want to come back with you after the comment period to get your feedback and

consideration of those comments.

MEMBER HOPKINS: So that report you referred to is going to encompass everything, not just the measures that we already voted on, but everything that we are looking at today?

DR. WINKLER: What we are going to do is put --

MEMBER HOPKINS: Just these?

DR. WINKLER: What we are going to do is put these out in two ways. A couple of reasons: It gets a group of them out forward and faster. It is a little easier on audiences to digest that number of measures. So we are releasing them as two publications.

You are right. Down the road where final-final comes together, they will get packaged together, but for right now --

MEMBER HOPKINS: The other question: Are we going to see that draft report before it goes out?

DR. WINKLER: Yes. We are

planning on sending it to you. We didn't think you wanted it like last Thursday when it was ready, with everything else. So I will be happy to send it to you tonight or tomorrow night. Absolutely, we will be sending it to you for your comments, but we didn't want to over -- You have plenty right now. We want to give you a chance to take a breather.

CO-CHAIR DUBOW: I just want to remind you that the measures that we don't recommend are also included in the four and subject to public comment as well. So that you can see, we will have the opportunity to gauge public response on all of the measures.

MEMBER HOPKINS: So the results that you just told us, what got voted for, what didn't, that came out of our voting survey. This is the first time we have heard what those results are.

DR. WINKLER: The reason it took a long time is some of you voted conditionally. You recommend with conditions. I had to sort

through all those conditions with the measure developers and figure out whether they responded and whether your vote ultimately became a yes or a no. That took a while. That takes a while. So, yes.

We also have -- I just finished them -- the summaries with the actual votes. Those will be available. Again another thing to circulate to you; didn't think you wanted it last week, and we will be sending those to you as well. they will be part of the information that is also posted. So it is in view for everyone on how the votes came out, not by individual but in the aggregate for the committee.

CO-CHAIR DUBOW: If there are no more member comments, we are now open to entertaining comments from the public.

MEMBER GIBBONS: I have a comment.

CO-CHAIR FLEISHER: Please identify yourself.

MEMBER GIBBONS: Ted Gibbons from

Seattle, Washington.

CO-CHAIR DUBOW: We can't hear you.

MEMBER GIBBONS: This is Ted Gibbons from Seattle, Washington.

CO-CHAIR DUBOW: Are you speaking on a speakerphone?

MEMBER GIBBONS: No. I am on a headset.

CO-CHAIR DUBOW: It is really breaking up. It is very difficult to hear you.

MEMBER GIBBONS: Okay. I am sorry. Why don't you go ahead then. I will forgo my comment.

CO-CHAIR DUBOW: Ted, again we are going to listen very carefully. Ted?

MEMBER GIBBONS: I will not comment. Thanks.

CO-CHAIR DUBOW: Okay.

DR. WEINER: Thank you. I am Dr. Weiner from SCAI, and I just want to raise

some concerns over the PCI readmission measures. We have been on record now almost four occasions opposing the current measures.

I think with now the passage of the Patient Protection Affordable Care Act, there are even new wrinkles that, I think, need to be considered as part of this measure.

In the Hospital Readmission Reduction Program, there are actually now penalties for readmissions to hospitals, and that represents in the first year approximately a one percent cut across the board for hospitals who fail to meet the measure.

DR. GALLAGHER: This is Rita Gallagher. We cannot hear you at all.

DR. HALL: This is Bruce Hall from the American College of Surgeons. Yes, we are not hearing anything on the call.

CO-CHAIR DUBOW: Okay, hold on a minute. We are trying to sort that out. Can you hear us? Guess not. We are going to try

something else. Just a minute. Thank you.

DR. WEINER: I will start again, just for the benefit of the folks on the phone.

I am Dr. Bonnie Weiner from SCAI. I am an interventional cardiologist, and we have been on record now on multiple occasions being opposed to the proposed PCI readmission measure, and I don't know that we should go through all of the details, but certainly, the concern evolves around attributable readmission.

Now with the passage of the Patient Protection and Affordable Care Act, there is even a bigger issue, I think, which reflects on the potential for penalties as part of the Readmission Reduction Program, which represents about a one percent across the board cut to hospitals who don't meet an approved measure that will unfairly penalize hospitals with cath labs as opposed to hospitals that are potentially not at risk for

that one percent risk, because they don't happen to have a cath lab.

So I think that is an unfair advantage -- unfair disadvantage to hospitals who are providing high quality care and high value and high technical types of care for cardiovascular patients.

It is also important to recognize that in the Act there are specific exclusions for readmissions that are unrelated to the prior discharge, and that has been one of our big concerns about the way the current measure is designed, that it does not attribute the readmissions to the PCI; and because this is a measure that is specifically termed related to the procedure as opposed to the disease state that the PCI is treating, it seems unreasonable to us to expect the PCI hospital physicians and system to be accountable for readmissions that are unrelated to complications of that procedure or the process of that procedure.

As we have looked at the data, there is a significant number of the readmissions within the 30 days after PCI that are not attributable to the PCI procedure itself, and we don't believe that that is being compensated for appropriately, the way the measure is defined.

Finally, if this were to go forward, we think that, first of all, the 30-day is the wrong window for PCI, because there is a lot of noise beyond the first seven to 14 days that, again, is mostly unrelated to the PCI procedure. But if it is going to go forward in its current form or any form, we think this should be a time-limited measure, because we think there is a lot to be learned about how to better define what is attributable and not attributable to the PCI procedure itself. Thank you.

DR. NEWCOMER: Just a question. You mentioned that there is a lot of admissions after the PCI not related to. What

were your criteria to deciding what was and what wasn't, and just give me a rough percentage of how many were not related.

DR. WEINER: Sure. I don't have the exact numbers in front of me, but it could be as much as 30 percent. A good example is somebody who has a PCI who then comes back in to get their hip replaced or has a routine screening colonoscopy, because it happens to be their time to get a yearly colonoscopy.

So there is a lot of those diagnoses that are captured within the 30 days. There is no question, if somebody has a breathing event after a PCI and he needs a colonoscopy or an EGD, that would be an attributable risk, I think, because of the drugs we use and all the things that go on around the PCI. But for the routine screening ones, we have no mechanism to really sort that out from a coding standpoint. You know, there's a lot of attributable risks.

MEMBER JOHNSON: I am a

gastroenterologist. We don't admit people for a colonoscopy or endoscopy. So they wouldn't even be on the screen.

DR. WEINER: Oh, I mean, for some of the states, that is true. But I think for a lot of the states, they do get captured because of billing issues as inpatient procedures.

CO-CHAIR FLEISHER: And also you would not be doing the rectal surgery within 30 days of a PCI. Where is your data from that you actually gave us that 30 percent unrelated? The measure specifications have been provided to you by CMS?

MR. HARDER: There should be the top 100 codes in a chart at the end.

CO-CHAIR FLEISHER: So you are saying it is from CMS data that you are --

MR. HARDER: Right.

CO-CHAIR DUBOW: Are there any other questions? Dr. Johnson, did you have a question?

MEMBER JOHNSON: Just a comment. We actually have very specific guidelines and consensus recommendations. So we never do a procedure, unless it is an emergency, within 30 days, as far as any type of screening or routine elective stuff; because Plavix is non-negotiative for 30 days. A non-drug alluding stent, 30 days you can stop it, but we don't do -- and I am a past President of the College, and I was involved in all these guidelines.

So I will tell you that that just seems to be a very, very minute scope of the patients you are talking about.

DR. WEINER: And again, there's 100 codes, and I happened to pick those couple off the top of my head. There are certainly others. There are dialysis codes that are used, and again we are not talking about somebody who has acute renal failure and needs dialysis because of a complication related to contrast.

We can't sort out the sort of chronic dialysis patient who gets sick for other reasons, you know, two weeks later, and winds up getting admitted, and the code of dialysis is used as part of that admission.

That is why we think there is so much noise in the measure. In order to attribute it to the PCI procedure itself, we need a lot more information, a lot more granularity about how those codes are being used, and to not subject the hospitals to the potential of the one percent penalty across the board for all their DRG reimbursements just because we are not very good at defining what those attributable risks are up front.

DR. GALLAGHER: This is Rita Gallagher again. I am on the measure development team. Would it be okay if I made a comment about why we approached it this way?

CO-CHAIR DUBOW: Please.

DR. GALLAGHER: Okay. The challenge -- You know, we didn't design it so

that every -- We didn't design to account for tightly related -- readmissions that were related to the procedure just by the AMI going into heart failure. The all cause relation measures inside the group of four don't take a narrow view.

Then the challenge is, I think, NQF has seen in other measures like APR DRG readmission measure where you try to look at pairs of inpatient diagnosis and procedures, and then a readmission diagnosis like heart failure, and then readmission, and is this related or not, not just to the procedure but to the hospital around the procedure.

So this is a measure that measures not just the procedure, but the person's experience of hospitalization and discharge, coordination of care, follow-up, medication reconciliation, and there are a lot of things around that care that may not be related to the interventional cardiologist's actual technical expertise in the procedure that do

affect the possibility a patient is readmitted, and we actually want to capture the, quote/unquote, "aspect readmissions" relative -- looking across the spectrum of hospitals and seeing which hospitals had relatively high readmission rates.

We know there are going to be some readmissions that are unavoidable. We just know that. We are not trying to get to sort identify preventable readmissions and then get those down to zero. We are looking at a relative performance of readmission for hospitalization and the follow-on care for the whole episode of the care of the patient, and that is why it does not try to parse out is this closely or tightly related to the actual execution of the procedure.

I just want to comment. I am not sure what the breakdown is for doing that, but that again is as relative performance relative to peers. It doesn't require or demand that only particular kinds of readmissions are

counted relative to the condition or procedure for which the patient was originally readmitted.

DR. WEINER: Just two comments. One is I think what you say is true if we were talking about disease state treatment, but we are not. This is labeled specifically as a PCI readmission measure, not an acute coronary syndrome, not an acute MI readmission. It is specifically targeted at the procedure, and to not make the readmission related to the procedure, if you are going to call it that, I think, is disingenuous.

The second thing is that the law specifically says that the readmissions need to be attributable to the prior admission. So to just sort of ignore that and say this is, you know, somehow relative, when we don't even know what the right benchmarks are at this point, I think, is again a reason to at most make this a time-limited measure so we can gain that information, and at worst something

that we should really go back to the drawing board and think about again.

MEMBER GIBBONS: Well, this is one of the previous TAP folks here today. Can I make a comment about that?

I actually saw this as the measure developer's perspective, and I think that, although I highly respect the intervention opinion on this, I think it is important, though, that it is disingenuous to say that this is only related to the PCI procedure.

This is related to the disease states in the PCI with all the results and the need for the PCI and the complex medical conditions that are associated with individuals who need PCI.

So this is actually looking at the global perspective of the medical care, is not meant to reduce the readmission rate to zero, but to look at patients that can be examined, reevaluated and reduced by an individual institution looking at their own experience.

So I think that to say that this is only related to the PCI operator is actually is actually a very small part. This is looking at the condition that results from the PCI and the management of the patients who have just had the PCI.

CO-CHAIR DUBOW: We heard that you support the position of the measure developer, and then missed a lot of what you said. If you could send us a quick overview, just so that the record is very clear, it would be very helpful. An e-mail to one of the staff would be great. We did get the thrust, but not the nuance.

MEMBER GIBBONS: I'm sorry. I am actually on a microphone, but --

CO-CHAIR DUBOW: Okay. Well, I'm sorry about the technical issue, but just to ensure that we have your view, it would be useful if you could just dash a quick e-mail to staff, just so that we can have it recorded properly.

Are there any other questions? Otherwise, we have another public comment.

MR. HARDER: Hi. My name is Joel Harder. I am the Director of Quality Initiatives.

I am asking the measure developer -- I know that Lein is here -- if she could step up and ask us -- and inform us if this should be used for the Hospital Reduction Program per the legislation, or not.

CO-CHAIR FLEISHER: Is that question relevant to our decision?

DR. HAN: Yes. I don't think we should comment on the health reform bill yet.

MR. HARDER: It will be relevant to the comment period by the public, and I would like the transcript to reflect upon this information.

CO-CHAIR DUBOW: But I am not sure that Lein is in a position to be responding on behalf of CMS. So I don't think it is a question that she is in a position to respond

to. So it is fair game if you want to raise it, but I don't think that we have anybody here who can respond to it.

These regulations haven't been written, and we just don't have the information yet.

DR. HAN: And we have to work on how we are going to carry out those things in the health care reform bill, and I don't think that we are right now ready to answer that question. We have to work on that first. But I do have two responses to what you have raised.

One is that this is a hospital level measure. So we are not specifically focusing on the physician or the surgeon who performs this procedure. It is the whole management of patients. It is the whole -- the system result of the measure. So I just want to remind you that it is a hospital level measure.

The other thing is that I think,

if you want to talk about the health care reform bill, we are talking about excess readmission here. So it is not shooting at zero, but it is excessive readmission here.

CO-CHAIR FLEISHER: So may I ask, in relation to your first comment, so is this really more of an episode of care. The initial comment here was about this is a procedure, but it really sounds like the procedure represents an episode of care. Is that how you are defining it?

DR. HAN: How do you define episode of care?

CO-CHAIR FLEISHER: In other words, is it really just somebody happens to have a PCI and, therefore, that episode of care really --

DR. HAN: Inside a hospital?

CO-CHAIR FLEISHER: Yes.

DR. HAN: I think it is even -- because readmission, we talk about 30 days. So you have that discharge, too. How do you -

- you know, the whole care after. I mean discharge planning, that kind of process.

I think, if that is the episode of care you are talking about, yes.

CO-CHAIR FLEISHER: Yes, I am. So it is just not about the actual procedure.

DR. HAN: It is not one thing. We are talking about from patient perspective, that you got hospital, and you were readmitted in 30 days. I think the whole package of the hospital. That is my point.

CO-CHAIR FLEISHER: Thank you for that clarification.

DR. HAN: Okay, thank you.

MEMBER PINDOLIA: I think that you just said what several of us have been talking about is quite important, that this is a person. So we talk about patient centered outcomes, not a 15 or a 30 or a 50-minute window of time in that patient's life, and ignore all of the rest of the things that are happening around them.

So I really also strongly support the idea that this is a patient centered measure, and the patient comes in with many things, and to ignore the fact that people choose what happens to that patient in the hospital, regardless of what the patient may think at that moment, so they choose and do the procedure. They choose to give them this medicine or that medicine.

I just think that we need to look at a patient centered measure, and it is -- You have hospital care of that patient, and you need to keep remembering that.

CO-CHAIR DUBOW: Okay. Are there any other public comments?

MR. HARDER: Yes. I would like to continue.

CO-CHAIR DUBOW: Please.

MR. HARDER: I just want to let the Steering Committee understand that this is also an inpatient and outpatient population, and that in the outpatient population the

patient comes in for this procedure and gets discharged sometimes the same day, is the trend right now, and that 30-day window is fairly long, in our view, for the cardiovascular related readmissions that we are very, very interested in, and this isn't Nurse Sky's. This was evaluated by the TAP PCI Registry Steering Committee, which is where the patient population is coming from.

We really feel that, you know, for this to be of best interest to the PCI patient population, these are issues that we want to see in this measure.

CO-CHAIR FLEISHER: Other comments?

CO-CHAIR DUBOW: Are there any comments from the public on the phone? Okay. Thank you very much. Thank you both.

CO-CHAIR FLEISHER: Any comments from the committee? No.

CO-CHAIR DUBOW: Thank you. So we are now going to do the diabetes measures?

The intro, right. Sorry.

DR. WINKLER: Joyce wants to stay on time. Got it.

Just to kind of start the day and the work we have ahead of us is there is another group of measures that we are going to be discussing in the next two days. There are 28 measures. All right?

They are in a variety of areas, as we have listed them out. We will be discussing them as we did on the phone, but this way we have a slightly different dynamic with being face to face.

Measure developers are here with us. So what we will do is, much as we did on the phone, discuss them. We have tried to bunch them into groups, but a lot of the agenda and the order of discussion has to do with availability of those developers and some of the logistics behind that.

So this is what we are going to do today. Next slide: What happens with this

outcome is similarly, but about a month to six weeks behind the first group, this will also -- we will write the draft report. We will share it with you. The comment period will be in July, the voting in September, and the Board endorsement in October.

So we want them about a month apart, so you know where this is going, but they will follow the same pathways.

One of the things that the Co-Chairs asked, briefly before we get started, was a review of the decisions this Committee made in October when we set sort of the planners for this project around what are outcomes.

Just as a reminder of what we included, there was discussion as these measures were evaluated by the various Technical Panels of what are outcomes? Is this really an outcome measure? Where are those boundaries?

If you recall, this group cast it

relatively broadly, and just as a review: Measures of patient function, symptom, health related quality of life were in, as well as intermediate clinical outcomes, the biochemical and physiologic, which we have seen some of each; patient experience is an outcome measure, as well as measures around knowledge, understanding, motivation and adherence or health behaviors.

The next: Service utilization as a proxy for outcome, I think, is for the readmission measures or ED visit measures or some of those sorts of measures fall into.

Then there is nonclinical -- or non-mortality, clinical morbidity associated with a disease control or condition, and then adverse events or complications are outcomes, and then sort of traditional mortality is, obviously, an outcome.

So these are the parameters that you all agreed on would be our definition or our scope of what the meaning of outcomes is.

So the question does come up in a couple of the measures, whether it fits under this group or not.

Some of the measures are composite measures, and they seem to have a mixture of process and outcome, and because of the outcome component, these measures are being considered under this project.

So it is within the purview of this Committee to determine that a measure just does not fit under the definition, if you so choose. The TAPs discussed some of these and offered their opinions as well for you to consider.

So are there any questions? This is a sort of reminder/follow-up before we start launching into things.

CO-CHAIR DUBOW: Just to keep this in mind and to keep this slide handy, because I think it is going to -- I think we are going to need to refer to it, to remind ourselves about the parameters that we set for the

definition of outcomes.

David, did you want to say something?

DR. HOPKINS: Yes. I think another way to look at this is we sort of set out for ourselves the concept of what we consider to be a full dashboard of outcome measures, and at some point in this meeting I hope we have the opportunity to sort of revisit how well we did at filling up the dashboard. That is probably at the end and not the beginning.

CO-CHAIR DUBOW: Right. And, Reva, please remind me. The report will include research recommendations. Is that right?

DR. WINKLER: Yes. The report includes any recommendations that you make that accompany it, and whether they are research or whatever. We had three recommendations with the last group of measures.

There is also the second part of this project which, because we need to get through the measures, we aren't going to spend as much time, but we will get back to, and that is on the gaps, filling the gaps in the kinds of measures.

We have been asking to the Technical Advisory Panels for their recommendations. We are using this as the framework, so that for each of the various conditions that have been considered, do we have a measure? If not, what kinds of measures would be desirable to fit into each of them?

That is sort of an ongoing process as we go through the rest of this, and it is another part that, at the very end of all of this, will get packaged together. But you will have an opportunity to weigh in and review that as well.

We are trying to break this down into digestible pieces, but it is still an

awful lot of information to manage right now. So while we will take notes on anything that you raise during the discussion, we will save the focus discussion on the gaps to a later time when we don't have this work to deal with.

CO-CHAIR DUBOW: I think there is another overarching issue that we will have to deal with, and that is the concept of best in class, which NQF supports. We have a few measures that appear to be duplicative, but not entirely. On occasion they use a different data source, for example.

So we are going to have to come to grips with, I think, the help of the staff to determine how to handle those measures. Do we endorse two measures that are seemingly very similar or do we make a decision about a preferable data source, for example? So I think we just need to keep that in mind, too. Barbara?

MEMBER YAWN: And I had a question

related to the data source for some of the measures before. They were based all on CMS data, and so when I voted, I made it conditional that they only be applied to CMS data. I am not sure that I understand, when the measure is released, what is said about it; because it does change the way I would vote on some of these.

If they are going to use something else than CMS data, I may be quite uncomfortable with some of these measures being applied.

DR. BURSTIN: This is an interesting discussion. We have had lots of discussions about this, since you raised it on our conference call. Thank you, David. And it is really an important question.

The issue at this point is the fact that these measures include risk models that have been explicitly done on the basis of patients over age 65 in that population. I think we strongly want to move toward getting

to the point where we have measures that allow us to, in fact, have a different risk model or a companion risk model, and we are trying to think through what those next steps would be.

I think the idea that you would take a risk model developed for an over 65 population and just assume it would work in an under 65 population, I think, is difficult. That is where --

I'm sorry. The second issue is just, for some of the populations like, for example, Medicare Advantage, I guess the issue more so is I think the risk model probably would still work, but the data availability, I think, becomes the bigger issue.

So that is our understanding of it. Any other discussion would be very welcome.

CO-CHAIR DUBOW: Just with respect to that, I thought we had, in certain cases, made a recommendation to the developer that the developer work on making the measure

applicable to the missing population.

MEMBER YAWN: And you got the responses from them.

DR. WINKLER: Right. At this point, they would agree that the approach can be applied to other populations, assuming you have the appropriate data, and that the risk model would have to be adjusted.

So they are not saying you can't use it, but it is not something you can pull off the shelf and just plug in for another population.

MEMBER HOPKINS: So I still have my question. Why can't the measure be represented as applying to a population of people over 65 years old, and not specific to a segment of the Medicare population?

MEMBER YAWN: I mean, I would have no problem with that. I just have a problem with saying this measure is approved, and it just seems like, okay, if you are 50 and you have this, it should also include you. That

bothers me a great deal. So that is why I said some of my conditional approvals.

CO-CHAIR FLEISHER: Sean, did you want to make a comment as someone who reviewed these measures from a methodologic standpoint?

DR. O'BRIEN: I would say that some of the qualities of the measure you have to consider have to do with the reliability of the data elements and the data capture, and it is hard to really define that just in general without reference to a specific population and a specific source of the data.

So I think, for creating scientific acceptability, it is helpful to define the source.

CO-CHAIR DUBOW: On the other hand, or in addition, we need to make a decision about whether we stand on principle and not recommend a measure that otherwise has great merit for that particular population, and I would remind you, David, that virtually all people over 65 are Medicare beneficiaries.

I think maybe under three percent aren't. So it is a pretty encompassing --

MEMBER HOPKINS: Everybody is not fee for service. That is my point.

CO-CHAIR DUBOW: Well, just fee for service. Yes, well, I think there are -- I mean, on one of the measures I have a question about that, too. I think that is fair. But just hearing you talk about Medicare versus everybody over 65 --

Okay. All right. So we need to make these decisions about how finely we cut this thing. But I think that CMS should receive this opinion from the Steering Committee, certainly. So I think that is legitimately included in the report.

MEMBER YAWN: But there may be groups other than CMS that choose to grade hospitals and other things on some of these, like the PCI measure. We have a whole lot of people less than 65 getting those measures now.

CO-CHAIR DUBOW: Well, that is the under 65 issue.

MEMBER YAWN: Yes, that is. That is what I say. I think there are two issues here. One is does it only apply to the CMS group that we have data for, and the other is can it apply to all ages? I think, yes, we have to separate them, but it seems to me that the age one -- You can't just ask a statistician about the age group, because it has to do with the medical condition.

CO-CHAIR DUBOW: We have made those distinctions, I think. I think the staff is clear on them.

DR. BURSTIN: I do think we ought to continue to follow up with CMS, because I think David's point about why you would exclude just to keep a service based on a data availability issue is not a methodologic concern. It is an issue of data availability. If the data was available for those other plans, I don't -- You know, the model will

specify for those under 65, not for those in fee for service. I agree. We will follow up with CMS on that.

CO-CHAIR FLEISHER: So one of my questions would be: If we endorse this model for the fee for service, do you have to actually go back to NQF to get it endorsed for Medicare Advantage or --

DR. BURSTIN: We will try to work those issues through before we put it out for comment.

CO-CHAIR FLEISHER: Great. Perfect.

MEMBER GERBIG: But is it correct that, once a measure is approved, it is available for any payer to use or can a measure be limited to only a 65 or older?

My understanding was, once it is approved, it is available to anyone to use.

DR. BURSTIN: This is, again, this e-mail exchange David and I had, that in some ways that almost becomes like an off-label use

for FDA. It is pretty analogous.

It hasn't been tested on that population. Do you really want to use that drug with the potential risks and issues involved in using it? I think, you know, if the measure is specified for over 65, I think those who choose to use it for under 65 could potentially have some issues on their hands, and I think it would not be wise, although I think we would like to work with the measure developers in general to bring forward measures that allow us to have risk models to get as broad a population as possible.

CO-CHAIR DUBOW: In other words, there is not an NQF police department, and you know, this is a voluntary process all around, maybe not at CMS, but it sort of is. But we won't go into that. But the point is that NQF specifies these measures and then trusts that they will be used appropriately.

CO-CHAIR FLEISHER: So to say, Helen, that somebody is using an NQF endorsed

measure, that would not be NQF endorsed.

Did we want to comment on best in class?

DR. BURSTIN: Best in class, we have to come to as get to those issues. I mean, I think the thing from our perspective always is evaluate the measure in front of you exactly as it should be, based on the criteria, and then after you have evaluated that measure, I think it is appropriate to look toward whatever is already endorsed and make a decision on whether it actually adds something to the portfolio or is it really sort of just -- using the FDA analogy, is it just another sort of "me, too," kind of --

CO-CHAIR DUBOW: And that is a criterion in the measure evaluation list. So we have that to assess as part of our work. Okay. Are we --

CO-CHAIR FLEISHER: We are on time. We are going to go into the diabetes measures. We are actually going to -- Reva is

going to go over the measures, but we are going to take them out of order in that we are going to do the single measure first, the HbA1c, and then we will do the two composite measures second.

DR. WINKLER: So Hawa is going to bring that up. All right.

CO-CHAIR DUBOW: Is the measure developer here, by the way?

DR. WINKLER: There he is.

DR. BURSTIN: Is anybody from Minnesota Community Measurement on the line? Dan? Anybody?

CO-CHAIR DUBOW: All right.

DR. WINKLER: Let me just -- I am just trying to find it. Okay.

NQF has endorsed a set of diabetes measures for pretty close to its entire existence. Within the group of endorsed measures, there are a large number of outcome measures. Hemoglobin A1c levels, blood pressure levels, LDL levels are measures that

are endorsed by NQF for years.

Currently within the portfolio we have, in terms of Hemoglobin A1c control outcome measures, we have endorsed the measure that is poor control, which is Hemoglobin A1c greater than 9, and most recently we have also endorsed the measure of Hemoglobin A1c less than 8.

This is another of sort of a set of measures from the same measure developer on Hemoglobin A1c outcome measures, and this is for patients 18 to 65 years of age with either Type I or Type II diabetes with a Hemoglobin A1c level less than or equal to 7 percent.

This measure focuses in on a narrower population than the other measures, the other outcome measures that we have endorsed. I think this one is one for selected populations.

So the numerator for this is the most recent Hemoglobin A1c level performed during the year of 7 percent. The applicable

population is aged 18 to 65 years. This is a younger population. The other diabetes measures apply to patients up through age 75.

This is a more aggressive management target, and the -- go ahead and scroll down, Hawa. It is not a process measure. It is an outcome measure. It is an intermediate outcome measure.

The Diabetes Technical Panel did review this measure, and they felt that again it was -- They rated it high on importance: Outcomes for diabetes, large population, getting them under good control. Appropriate intermediate outcomes do reflect long term outcomes.

So the only issue was, because this is a narrower population and it is an aggressive outcome target, has the population been managed sufficiently enough to be appropriate for that lower level and more aggressive target? So that was rated highly.

On the scientific acceptability,

they rated it generally highly again. The measure is based on administrative data. They were concerned that there might be some exclusions not addressed, particularly patients experiencing frequent hypoglycemic episodes because of the aggressive target, people who have occupational risks for which you wouldn't want to have them experience those episodes, patients who are already on multiple medications and kind of maxed out on treatment and are realizing that for this measure not doing the A1c does count against. So it is not -- patients who haven't had the test done are included in the numerator.

They again rated the usability of this measure highly. It is a straightforward intermediate outcome measure similar to the others, and feasibility is good. These measures are already in use and use the same methodology.

So that is the measure before you, and I don't think Dr. Greenfield is here as

the TAP Chair. Is he?

DR. BURSTIN: Shelley, are you on the line?

MEMBER PINDOLIA: Reva, I have a question. I still don't quite understand what is the difference between this measure and the current NCQA HEDIS measure for HbA1c less than 7? It is just the age?

DR. WINKLER: Ben, correct me if I am wrong, but age is the primary difference.

MEMBER PINDOLIA: So it is 18 to 75 for the current one, and this is 18 to 65?

DR. WINKLER: That is correct.

MEMBER PINDOLIA: And that is the only difference?

MR. HAMLIN: This measure was just recently revised, actually. We have been collecting the less than 7 for a couple of years now. We added the additional age exclusion plus some cardiovascular and other comorbid exclusions as well for this particular population, given the new studies

in the coordinated events trial.

So we further restrict this with management exclusions for cardiovascular patients. I don't have the exact list in front of me, but it restricts the age as well. It is trying to target the younger, healthier population, particularly, for active management.

MEMBER JUSTER: The previous measure was 7 or 8, because I thought it was eight?

MR. HAMLIN: The previous measure actually was 7. It was not restricted. The 8 was just recently introduced last year during the review when we also further restricted the 7 population.

MEMBER JUSTER: And, Ben, two questions for you. Does the current NCQA measure -- is it less than or equal to or is it less than?

Second, are the regular HEDIS measures -- If somebody didn't have the test,

are they considered the same as a person who failed the -- In other words, they are both not numerator?

MR. HAMLIN: Yes. If the value is not present, they are included in the numerator, but they don't get credit for the numerator. So they get dinged.

I believe it should be less than or equal to 7.

MEMBER JUSTER: Okay.

CO-CHAIR FLEISHER: Barbara?

MEMBER TURNER: I'm curious what the unit of analysis is here. Is it a plan or is it a provider's panel, and are they looking at the main controls, so it is okay that you have -- you know, expect a certain number of outliers that are going to be in your panel?

MR. HAMLIN: This is applied to both the provider population and in the health plan population for both a diabetes recognition program providers and for the health plan.

The unit of analysis is the last measurement taken during the measurement year, which is our 12-month period from January to December. We have done testing in the past for these lab values of those relevant to, if you will, the most reliable data for that measurement period. If it was done multiple times over the year, it would not deviate from the mean. Generally, the last value is as close as you are going to come.

We do have further -- The approach is all in retrospective claims based approach with the health plan population. So --

MEMBER TURNER: Right. I just have a follow-up question. Do you have a minimum end per provider that you would insist on having to be able to have stable estimates, and have you looked at or thought about a change, so if you have somehow a high risk population, most of them coming in with A1c's of 10, you get credit for getting them down below 8 as opposed to getting dinged because

they are at 9?

MR. HAMLIN: Right. Yes. For the Diabetes Recognition Program, there is a minimum of 25 patients that meet these criteria, and this also -- It is important to understand that in both of these programs, this is one of the three HbA1c measures. So you look at the greater than 9, the less than 8, and the less than 7, and you get credit for wherever your patient falls in that population.

It is more the proportion of patients we expect to see below 7 versus below 8 versus above 9, and the Diabetes Recognition Program weighting is skewed to that fact, as is the performance score for the health plan population.

So we do expect to see -- And basically, we expect to see a certain proportion to fall within these certain parameters, and the weighting for the Diabetes Recognition Program takes that into account.

CO-CHAIR FLEISHER: David?

MEMBER JOHNSON: The question, I guess, is just in the unforeseen consequences, and I am not familiar with the Hemoglobin A1c as a gastroenterologist routine measurements. That is not what we do. But what happens to the people that are poorly compliant in situations where the doctors taking care of them have the staff, and the patient's understanding and the educational levels? I could see a drift away of avoiding poorly compliant patients just so you don't get pulled out as a bad provider here of this measure.

MR. HAMLIN: Right. I am sure that there is a certain amount of gaming of the system, if you will, for selecting populations. But in general, since we are looking at this primarily in the health plan population, you know, with 7 million providers and even more millions of patients, it is one of the factors of life that we have to take.

We are basically looking for the overall population. This is reported as a reasonable rate by plan. So it is the entire plan population; and in the Diabetes Recognition Program, you know, you have a small end population, you can select patients.

It is a continuous selection process. You are supposed to select 25 charts with the criteria, remove the ones that are not appropriate, select an additional 25, and go through that process. You know, I can't control the way providers select their charts for patient reporting in that program at this point.

So there are certain assumptions that there may be some selection bias there as well.

CO-CHAIR FLEISHER: Lee?

MEMBER NEWCOMER: I just want to -- Kind of reading through the measure, this looks only at patients less than 7, and your comment about range between 7 and 8, and 9 and

above, there are other measures that do that. This one does not.

I just want to quality. This stands alone at less than 7 only?

MR. HAMLIN: Right.

MEMBER NEWCOMER: Because we will be looking at one shortly.

MR. HAMLIN: Right. The greater than 9 and the less than 8 have already been endorsed as individual measures themselves. So we are looking to add to both this as an individual measure of less than 7 as well as a component of our composite measure for diabetes.

CO-CHAIR FLEISHER: So, in fact, this is very similar to some of the previous measures in that we could endorse this independently or as a part of a composite.

MR. HAMLIN: We prefer the former.

CO-CHAIR DUBOW: This is right now being considered as a standalone measure.

CO-CHAIR FLEISHER: David?

MEMBER HOPKINS: I don't appreciate looking at this measure as a standalone, because I think it absolutely goes with the other two.

Then I have a question for NCQA and NQF. Does this replace the existing endorsed less than 7 measure?

DR. BURSTIN: There is no endorsed measure. It has never been endorsed.

MR. HAMLIN: We are using HEDIS for several years. Now we are seeking endorsement with this new refined measure.

MEMBER HOPKINS: Well, I think it is part of a suite with the 7, 8 and 9.

DR. WINKLER: You could have the option of recommending it that way. Recommend that the measure is going to be an independent measure, but it should be used with the other two measures, because you do get a better holistic view of the population.

MEMBER JOHNSON: What would be the rationale for keeping it standalone?

MR. HAMLIN: You know, really, I don't know as a full standalone measure to really speak to that. We generally use it combined with the other HbA1c measures. I would imagine that, if a provider wished to use this measure as a standalone measure just to understand what proportion of the population was under 7, if they were doing a different program -- We have our own programs. We know that a number of our measures are used in other programs, particularly with the ones that NQF endorsed.

I do see value in keeping a lower level for HbA1c target, if you will, in the younger and healthier population. Generally, if the measure doesn't get endorsed as a standalone, it is harder to make justification for inclusion in the composite use.

All of our other measures that are currently in the composite are NQF endorsed as individual indicators.

CO-CHAIR FLEISHER: Dianne?

MEMBER JEWELL: You have already answered my question. Thank you.

CO-CHAIR FLEISHER: Barbara?

MEMBER YAWN: And this is 65 and younger. Is that correct?

MR. HAMLIN: Eighteen to 65, yes.

MEMBER YAWN: And the average person at 65 has how many chronic conditions?

MR. HAMLIN: I couldn't speak to that right now in particular.

MEMBER YAWN: About three on average. So do you really -- You know, I believe 65 is young. I am not sure I believe it is unhealthy. I am concerned about the age, of 7, and the data that is coming out about the side effects and the problems we are causing between ages 50 and 65 to patients who we are trying to push down to 7.

So I am concerned about the upper age limit of this, and perhaps you could tell us what the data is on the upper age limits and the risks to those people.

MR. HAMLIN: Yes. Actually, you know, the upper age of 65 was -- I mean, despite the fact that that population is actually getting younger and healthier every year, it was selected because of the fact that we do a retrospective claims based approach for our measures.

The DOPSI, you know, in the primary program, the health plan HEDIS population, we are looking at a retrospective approach. So we have to draw certain parameters around the population. The reduction of range to get it into the commercial Medicaid only, we have two product lines that we collect for this measure.

That was why that age limit was selected. The additional comorbids that we added, the cardiovascular disease and other comorbid conditions that also excluded patients in this population, would suggest your other point of the ones who are not healthy at 65, would also be removed from the

measure.

CO-CHAIR FLEISHER: We have that on the screen, just in case anyone is looking.

MEMBER YAWN: If I could see that.

DR. BURSTIN: I'm sorry. It is also on your thumb drive. I think it is page 55 of the diabetes -- Slide 60 of the diabetes risk file.

I just want to point out that, although the denominator is up to age 65, there are a very large number of exclusions to specifically get at the comorbidities, and that figure should speak to the specifics of the comorbidities rather than the generalities.

MEMBER YAWN: Yes. And that is great. I saw the cardiovascular.

MR. HAMLIN: Right. We have chronic renal failure, dementia, and the other ones you will see there as well. We are also looking at this -- I mean, again, this, of course, we have taken a retrospective claims

approach primarily.

We are looking now to further refining these definitions through, obviously, electronic health record environment and new coding that are going to be available, but right now with the approach that we have and the data we have collected in the last couple of years for this measure, these seem to be -- and our expert panels and the TAP agreed that this was sort of a reasonable comorbid list to include. This has all been through codes as well. So there are complete code lists.

CO-CHAIR FLEISHER: I think B.J. was next, and then Amy.

MEMBER TURNER: Thanks. So I think pretty much the evidence of the value of pushing A1c down to this level is for people who have an estimated survival of at least 10 years to 20 years.

There are lots of people who have those conditions that are not exclusions, like cancer, etcetera, and their list would have to

be much more comprehensive, and it would actually have to be much more evidence based. This seems like a somewhat random selection of folks with, say, cardiovascular risk.

So please explain to me why you need to apply this across the board without having a much better sense of what someone's estimated life span is.

MR. HAMLIN: Well, again it was the -- You know, the criteria that are outlined here were the sort of best judgment from the expert panel and from us on the appropriate exclusions for this population, given the evidence for driving A1c.

Obviously, the folks who are primarily who are on microvascular and macrovascular with long term complications for driving A1c down -- you know, I can't speak to the cancer at all at this point. I don't have the background to provide you there.

MEMBER TURNER: That is the point, I think.

MR. HAMLIN: Right.

MEMBER TURNER: I mean, this is the thing. It is very -- well, myopic is a bad word to use here. But it is very limited, and the evidence right now, the adverse consequences of pushing somebody down to that level, is just emerging right now.

You have two perfectly reasonable measures, the 9 and the 8, to be able to get people to improve their care and continue to strive, but to push people to a level where you don't really know there is a benefit for them, because they may not be living that long, and they may certainly be getting side effects that you can't capture with the kind of claims data that you use, seems to be having significant unintended consequences.

MR. HAMLIN: Well, you know, I would disagree. There is actually evidence to show that moving certain patient populations who have a diagnosis of any kind of cancer may be inappropriate. I do agree that there are

probably specific diagnosis of cancer or perhaps current treatment regimens that might want to exclude them from the population, but I don't believe the evidence shows that someone who has a diagnosis of breast cancer that is, you know, Stage 1 should not be managed well and actively if they are also diabetic.

MEMBER NEWCOMER: There is evidence to look at age and look at complication rates for hypoglycemia as you age. Is there any evidence on that?

MR. HAMLIN: Interestingly, when we first -- When we filtered this data and when we collected the first year data, we had forgotten to address certain restrictions on the age around our patient population. When the data came in on the first two years, actually the over 65 -- the 65 to 75 population actually had better A1c rates than the other two populations combined.

So it was one of those areas where

we still feel like it is an inappropriate population to measure through our approach. However --

MEMBER NEWCOMER: Well, what is the evidence in the medical literature? What we are trying to drive to here is we know that hypoglycemia is a definite consequence of this kind of tight control. Does that vary from age population to age population in the literature?

For instance, can a 20 to 30-year-old tolerate that better than a 50 to 65-year-old? Do we know?

MR. HAMLIN: We don't know the exact parameters around what ages would tolerate the aggressive management better than others at this point, but again the experts felt that 65 was a reasonable cutoff for active management, given the new trials that just came out.

MEMBER NEWCOMER: So we are talking about an opinion versus any evidence.

That is all we are trying to clarify here.

MR. HAMLIN: Yes.

MEMBER NEWCOMER: Okay.

CO-CHAIR FLEISHER: Okay. Amy?

MEMBER ROSEN: I just wanted to follow up on that, that this measure is not risk adjusted, and given the concerns, I am wondering why risk adjustment was not considered, and something like age and gender could easily be folded into some type of measure like this. If not, risk adjusted, at least some stratification could be done.

I also had some concern about the denominator in that you are using claims data, but oftentimes somebody may come in with multiple problems, and a diagnosis of diabetes might not get on the claim. So that won't be in the denominator.

I wondered if you had thought about looking at pharmaceutical claims as a way of getting a more comprehensive denominator, and also had you looked at the

reliability of the medical record review or the automated laboratory data.

I know from my experience that, depending on where you look in the medical record, you may get different values of a particular test. So I just wondered what kind of guidelines there are in looking for that particular numerator that people might follow, so that you get a consistent reading from all the different providers.

MR. HAMLIN: To your first point, we don't actually risk adjust any of our effectiveness of care measures. We only risk adjust currently our cost of care measures for HEDIS. That is being examined right now, whether we need additional risk adjustment strata applied to more of our effectiveness of care measures, but that is a long way off at this point.

We do validate and verify our data collection methodologies. So before a measure can make it into the HEDIS population, we do

do both a claims and a medical record review validation study.

So we have looked at the reliability of the claims against the medical record and what turns up in the medical record versus what is available through electronic claims.

To your third point, we do actually include pharmacy as an identifier for diabetes denominator. It is a visit with a diagnosis or a number of ambulatory prescriptions for anti-diabetic agents. So that is included as part of the identification criteria.

CO-CHAIR DUBOW: The stratification issue?

MR. HAMLIN: The stratification -- We don't risk adjust or stratify the majority of our -- We don't risk adjust any of our HEDIS measures. We don't stratify the majority of our measures at this point, but it is something that we are looking into as we

move forward, whether that is something that we should be asking additional data elements from the plans.

CO-CHAIR DUBOW: But you do stratify in a way by payer.

MR. HAMLIN: Yes. We do stratify a commercial Medicare product, yes.

MEMBER ROSEN: That is not what I am suggesting.

CO-CHAIR DUBOW: I know, but it is a form of stratification.

MR. HAMLIN: Right. We do report all these three separately. So the commercial population, the Medicare population, the Medicaid population all are reported separately. For this one, it would just be commercial and Medicaid.

CO-CHAIR FLEISHER: Do we have any new topics that need to be covered?

MEMBER AMARASINGHAM: I just have one question. If a person -- If a provider is trying to get a patient below 7 and the

patient was experiencing hypoglycemic effects, how is that accounted for or is the patient excluded? I didn't see that in exclusions.

MR. HAMLIN: It is not at this time. If a provider is -- You know, again, these measures are guidance. They are not absolute. We don't expect a provider to go against the best clinical practices for managing individual patients.

This is a whole population approach that we are looking at. So, you know, we are not trying to tell physicians they have to manage them down to this certain level. They have to use their better judgment.

CO-CHAIR FLEISHER: Vanita?

MEMBER PINDOLIA: Hi. I just had one comment, again to have NCQA consider combining these with a Hemoglobin A1c 8 and 9 measure. The reason is that, just looking at state of Michigan and, I am sure, other states, how the HEDIS measures are being used

-- the five large HMOs, they are really just going after those less than 7, even though it is not NQF endorsed.

The Blue Cross/Blue Shield Michigan PGIP program for physician incentive only targets if you got them less than 7. They don't even consider the 8 or 9. So there seems to be a misconception of less than 7 is good for everybody, and by putting them all as individual, they are getting to pick and choose, and it might lead to some major patient negative outcomes.

MEMBER AMARASINGHAM: I would like to underscore that. I also think that, even among providers and plans, I think there is variation in sort of the baseline Hemoglobin A1c in the population. So if a plan or a provider had a higher proportion sort of in the 8, 9 range, I think this measure could be interpreted as getting everybody down to 7, which could have a lot of unintended consequences.

So I think it has to be a suite of measures.

CO-CHAIR DUBOW: Excuse me. If somebody on the phone has -- Everybody on the phone should be on Mute.

CO-CHAIR FLEISHER: David, a new topic?

MEMBER HOPKINS: No, no. I was actually going to move approval of this measure as part of a suite that would comprise less than 7, 8 and 9, less than or equal to, I think, and that is not a composite, by the way, if I understand the term composite.

So we are using a different term here, which is suite, which means it is three distinct reported measurements that are reported together.

CO-CHAIR FLEISHER: So we actually need David to comment, because that has not been proposed. All that has been proposed is your suite of pairing in a composite.

DR. WINKLER: Well, actually, this

is the purview of the Steering Committee, to do what has been done many times in the past at NQF. That is, we called them pairs when there were two. I don't know what you want to call them when it is three, you know, whatever. But the concept is not a new one. It is an old one, and it is independent of how the measure was developed or might be used.

If you feel that your recommendation for endorsement, that these should go together as a whatever you want to call it group, suite -- and what we do is it will be put out for comment that way. We will tag the other two measures to it.

When we put it out for vote, it is this and this and this, and you are voting on it as a group, so that they are an entity that rises and falls together.

MEMBER HOPKINS: That is my motion.

CO-CHAIR FLEISHER: Quick comment, B.J.?

MEMBER TURNER: So we have to respond to that motion, up or down vote?

CO-CHAIR FLEISHER: Well, do we want to go to public comment first? Are there any public comments before we move for a vote?

MR. HALL: This is Bruce Hall from the American College of Surgeons. I have a couple of questions.

MEMBER NEWCOMER: You are not entertaining the motion. Is that right?

CO-CHAIR FLEISHER: Not quite yet. Who on the phone is speaking?

MR. HALL: It is Bruce Hall, American College of Surgeons.

CO-CHAIR FLEISHER: Do you want to comment first?

MR. HALL: I have a couple of questions for the developer essentially on reliability and feasibility.

I saw that they provided a sample set of calculations that was roughly 550 patients. I was wondering if they had any

sense of how many practices can meet that sample size? I did not see any other commentary on the reliability of the patients between providers. So I was wondering if there is any information on that.

Then on the topic of feasibility, I see that they have made an estimate of how many hours of data collection would be required from both administrative folks and the medical record review.

I was wondering if there has been any discussion of the cost and the burden that is created.

MR. HAMLIN: I have to make one correction. The sample size of 550 is a recommended sample size for the medical record review approach. This is a paired measure we call a hybrid approach where you use a -- you select a sample, and do medical record review to collect the data elements such as HbA1c levels where they are not available just directly in claims data.

Our normal for the other measures is 411, but because of the additional exclusions applied to this measure, we had to up that sample size. That is applied to plans only, not to medical -- individual providers or medical groups.

The provider sample size is 25, as we said before. So there is a difference in the two methodologies for which product each measure is used in.

As far as the cost burden, again as we collect the data through this hybrid methodology of claims versus the medical record approach, we annually review the variation and performance among plans that are reporting as admin only -- the integrated delivery systems generally have this through electronic data -- versus the medical record approach, and we look at the rates, the variation of rates among the plans and then the variation of rates plan year to year and try and move that to a claims based approach

whenever possible.

Unfortunately, this is not one that is possible at this current time.

CO-CHAIR FLEISHER: Okay. Other comments?

DR. JEWELL: Hi. I am Kay Jewell. I am a physician, a consultant in Wisconsin, and I have two disclaimers. One is I am a consultant with one of the drug companies that has diabetes related products, but I also have a conflict -- well, influence in some work I do with consumers.

I have two points relative to the selection and the exclusion of the age 65. If that was selected, as I understand, a couple of years ago immediately after advance and reported and published, it was a reaction to concern of unintended consequences, and it has not been reviewed, and the actual evidence for excluding a 65-year-old person who doesn't have comorbidities from achieving a A1c of less than 7 -- I don't believe that there is

evidence for that as a risk factor and is an issue for hypoglycemia.

In fact, just like your data of the over 65, that they are actually doing better at getting A1c's, and we aren't seeing large numbers of problems that have been reported; and in fact, the NHANES data from 2003-4 -- it is the less than 65-year-olds. Sixty-eight percent of them are achieving less than 7 percent in 2003-4 versus 48 percent for those less than 65.

So the data would suggest that the elderly are not having a problem with this, especially if you have a way to exclude the comorbidities.

The other concern: There also has to be a way for the physician to be doing the individual assessments. The ADA and the Endocrine Society do recommend an individual assessment, and looking at the comorbidities, and do not use age all alone as a criteria for not achieving good control, and there has to

be some room for individual physician assessment.

Hypoglycemia, as far as the data that I have looked at, what they have identified -- and this did come up at the TAP; Dr. Hellman talked about it, that it wasn't age per se that was the issue for hypoglycemic events. It was things like frequency of testing and attention to testing, and hypoglycemic awareness, which is probably different.

That is one of the issues and concerns about including this population, is that we are going to have all this hypoglycemia, because it is an age issue. I don't think the evidence is there.

In terms of looking at the long term effects, there are two issues of long term effects that come at 15 to 20 years for cardiovascular, but there are also short term effects for a 65-year-old.

There are short term effects in

terms of retinopathy, and eye disease can be achieved within five to eight years. There is also risks in terms of -- impact in terms of infection, infection control, and neuropathy. Neuropathy is an early issue, and a very, very important patient outcome pain issue.

MEMBER NEWCOMER: How do those differ from a population at 8 or less?

CO-CHAIR FLEISHER: I think, actually, we need to move on, because we have multiple measures. Unless anyone feels strongly, I would just like to comment that the TAP endorsed this measure as a standalone measure. Would you like to comment, Reva, on the TAP?

DR. WINKLER: In general, they felt that this measure narrowed a population. There were some more questions that you raising with it. Was it narrowed appropriately and enough to worry about the adverse consequences, but you know, they generally supported the measure.

CO-CHAIR FLEISHER: So I have actually heard a motion. I assume Lee was going to second it.

MEMBER NEWCOMER: I wasn't, actually.

CO-CHAIR FLEISHER: You weren't? Well, what I have heard is several either endorse the measure standalone, endorse the measure as part of a triplet, endorse it as part of a composite, or not endorse it. So I wanted to defer to Reva to see how we will vote.

DR. WINKLER: I would recommend voting each of those independently, and we will see where it leads us. Is there a logical -- You know, do you feel it should be a standalone measure, yes or no? Then the next one: Do you feel -- Would you recommend it as part of the three-group, yes or no? That way, rather than split the committee.

CO-CHAIR FLEISHER: And by hands? So I guess it is a show of hands with regard

to the committee, with regard to a standalone measure. All those in favor of a standalone measure.

MEMBER NEWCOMER: I need some help before we do that. One question on the standalone measure. Do we have evidence that the hemoglobin elastin-7 has -- that we understand what its long term side effects are for all populations? Was there evidence talked about that in the TAP?

I understand for certain populations, it clearly benefits. That is indisputable.

DR. WINKLER: They certainly discussed it. It varied between some evidence and a lot of opinion.

DR. NEWCOMER: So just tell me about the sum evidence for this large a population.

DR. WINKLER: I can't speak to the details.

DR. NEWCOMER: Okay. Thanks.

DR. BURSTIN: I would just refer you again to the TAP summary. They really did spend quite a bit of time on this measure, and they specifically felt that in general the evidence wasn't there for Hemoglobin A1c less than 7 for all patients, but they did specifically indicate, at least in what they said, that they thought that the no risk adjustment beyond exclusions was okay, and they specifically wanted to be sure that Stage 4 and 5 CKD was out, which it is.

So that is all we can share, unfortunately. I am not sure if Shelley has joined us yet on the phone.

CO-CHAIR FLEISHER: Before we vote, David, do you want to introduce yourself and any disclosures?

MEMBER HERMAN: My name is David Herman. I am from the Mayo Clinic. No disclosures.

CO-CHAIR FLEISHER: So I guess we are calling the vote. All those who would

like to endorse this measure as a standalone measure, please raise your hands.

Abstentions? No.

DR. WINKLER: No? That is everybody?

CO-CHAIR FLEISHER: One abstention. Okay. Second vote: As part of a group of measures, including 7, 8 and 9.

MEMBER TURNER: Point of clarification. I don't know what that means. I don't understand how you operationalize it. Does it mean 25 percent have to be under 7, and 50 percent have to be under -- How do you do that? I am just wondering how you operationalize it.

DR. BURSTIN: I think there is really two issues. I think what you have just said is on its own, you wouldn't, for example, want public reporting of Hemoglobin A1c less than 7 for selected populations on its own.

You then have an opportunity when we get to the composite to say, okay, maybe

not on its own, but maybe in the composite. I think the third issue that was brought up was that, since we have already endorsed the less than 8 and the greater than 9, I guess one other possibility would be to indicate your support potentially for Hemoglobin A1c less than 7 should only be publicly reported with the other two levels. That is what, I think, was --

MEMBER TURNER: It doesn't help me understand what it means. In other words, is it the goal to have 90 percent of your population under 7? Well, then how do you -- You said it was weighted, and I am trying to understand what weighting means. No?

DR. BURSTIN: I think there is some confusion between the Diabetes Recognition Program, which is not on the table, which is weighted and scored. Some of that will come up during the composite discussion, because I think some of the weighting is pretty similar to what we were

literally just talking about in terms of publicly reporting the measure.

Would you want to see the rates of those three levels for a given practice or whatever the case may be publicly reported, even if you didn't feel comfortable seeing A1c less than 7 on its own. That is all.

MEMBER NEWCOMER: Is this measure actually that, and wouldn't we vote on that measure alone, because the next measure does ask for 8, 9 and 7.

CO-CHAIR FLEISHER: But it includes multiple other criteria.

MEMBER NEWCOMER: So we are voting on would we like to see 7, 8, 9 standing alone?

CO-CHAIR FLEISHER: Correct.

MEMBER AMARASINGHAM: Here is my -- With respect to this suite of measures, sort of representing all three, what is the implication for the provider, though?

For example, if the provider's

baseline was, before they even saw the patient, 40 percent had a Hemoglobin A1c greater than 9, then what would it mean in the subsequent year to represent the proportions; because you don't have an anchor period.

So I am just trying to understand what is the purpose of the reporting? Is it to kind of reflect the provider's performance or to say this is sort of the baseline population?

DR. BURSTIN: It is intended to reflect the population.

MEMBER NEWCOMER: I would answer that as yes. The first year would be baseline performance. The subsequent years would be performance. So it is just a measure. All we are doing today is saying this is an acceptable measurement, and how it is used is going to vary from person to person.

CO-CHAIR DUBOW: I think we need to be clear that NCQA has to go along with -- If this is a recommendation to NCQA -- It was

submitted as a standalone measure. So we need to -- This vote is going to reflect the recommendation.

MR. HAMLIN: The reason it is a standalone measure is because the other two are already endorsed. It is the one left over.

MEMBER YAWN: The only advantage I see of it being part of the suite is you get to exclude some people. Now do I think it is the right ones, and do I have any evidence? No. That is the only thing that I can see that is positive. Rather than just saying, okay, we are going to do 7, 8 and 9, and have no exclusions for 7, this gives you some exclusions for 7, and that is the only advantage I can see of thinking about putting it out there.

Doesn't mean I am going to vote for it. I am just saying I think that is what we have to think about.

CO-CHAIR FLEISHER: Quickly,

because we do need to move on.

MEMBER PINDOLIA: I think the other advantage is how it is used out in the public. Once NCQA approves, the HEDIS measures are used to give payer incentives. They are used to have measurements of this health system is better than this one, and to have a trio, at least you can see overall, and you are not being forced to draw everyone less than 7. So there is another use.

MR. HAMLIN: My only one concern with putting it in a trio is that the greater than 9 and less than 8 do not have these exclusions applied. So they apply to different populations.

So there is some variance in those populations. The population of the greater than 9 and less than 8 are 18 to 75 with no additional comorbidity exclusions. The less than 7 is. So --

DR. BURSTIN: Just to be clear, just a point of process, you have already made

the initial assessment of should it be a standalone measure. No. Your only options at this point are potentially to put this back to NCQA with a recommendation or even a condition saying the only way the Steering Committee would approve A1c is if it was always reported with 8 and 9.

That is really your only opportunity at this point. They haven't submitted to you a suite of 7, 8 and 9. So at this point it would be your recommendation back to them indicating less -- and this is just a possibility -- less than 7 would be fine as long as it is reported with the other two levels. That could be a condition or recommendation.

CO-CHAIR FLEISHER: David, would it be acceptable, can we defer that vote until we talk about the composite? Do you want to vote now?

MEMBER HOPKINS: Well, there is a different problem, because you've got two

composites, and I believe they are in competition with each other.

CO-CHAIR DUBOW: So let's wait on that.

CO-CHAIR FLEISHER: So the vote on the second motion which Helen described of going back to NCQA and actually suggesting this be developed into a measure with all three. How many vote -- No?

DR. BURSTIN: No, sorry. Don't want to confuse things. They wouldn't actually develop anything. I think the issue would be this Steering Committee has indicated not as a standalone. The Steering Committee could potentially put forward a motion saying you would conditionally approve Hemoglobin A1c only if it is always reported with the other two levels.

It is just a reporting issue. It is not combining it into a composite, which is the later discussion.

CO-CHAIR FLEISHER: So how many

vote yes for that?

DR. WINKLER: Fifteen.

CO-CHAIR FLEISHER: Noes? So we have four noes. Abstentions? Okay.

MS. BOSSLEY: Barbara walked out of the room.

CO-CHAIR DUBOW: Oh, okay. If Barbara walked out, we didn't get her vote. Okay.

CO-CHAIR FLEISHER: Okay. Anybody else not vote? Well -- Do people want five minutes?

CO-CHAIR DUBOW: We have to finish this.

DR. WINKLER: The decision was to -- The next measure will be the composite measure submitted by NCQA, which is the comprehensive diabetes care measure. This is measure 29, and this is essentially a composite measure, the percentages measures.

This is 18 through 75 with diabetes related to the following, and you can

see the components of the composite. Ben, correct me if I am wrong. The most recent version I saw of this did not have two blood pressure controls. Correct?

MR. HAMLIN: Yes. If you downloaded this from your online site, we were having technical issues getting the updates in there. So, yes, it does not include 130 over 80. We were having trouble saving our changes after the TAP.

CO-CHAIR FLEISHER: So what is the current one? The current measures includes?

MR. HAMLIN: The current one measure includes everything you see here except for the less than 130 over 80, because the TAP asked us if we would remove that from consideration.

CO-CHAIR FLEISHER: And keep just the one?

MR. HAMLIN: Keep the one, 140 over 90, yes.

CO-CHAIR FLEISHER: Great. Thank

you.

DR. WINKLER: This is a composite measure that includes these components, all of these components with the exception of the Hemoglobin A1c less than 7 currently endorsed by NQF, and also the blood pressure control less than 130 over 80, which is not included in the most recent version.

What we need to go to, Hawa, is the table that talks about the weightings and how this composite is put together.

MEMBER NEWCOMER: There is an asterisk behind the Hemoglobin less than 7. Is that because it is not in the current NCQA? Is that what the asterisk stands for?

DR. WINKLER: No. Because it is not currently NQF endorsed, and your actions have relevance to what is going on here.

MEMBER NEWCOMER: Okay, thanks. So do you have that one, Helen?

DR. BURSTIN: Right.

DR. WINKLER: Okay. One of the

things, when this was initially presented to the TAP is these weightings were not available, and so they re-met by conference call last Thursday to take a look at how these measures are combined.

These are the criteria as well as the points given. This is the table for the recognition program. So the meeting 75, I think, is more an implementation issue, but the methodology for combining each of the -- all of these measures is really a two-step methodology; whereas, you get credit if you meet the criteria, and then you get that many points, and then the sum of the points.

So the final score is your total number of points. How a user or implementer might then use those points to display or publicly report, I think -- NCQA does it one way, which isn't necessarily the only way that it might be done, and that was the discussion on the TAP.

Again, as you will find with most

of the composite measures that we will be discussing over the next couple of days, the weightings and the choice of how many points and things like that are somewhat arbitrary, but based on the developer's value system around what is important in the care of patients around this condition with their Technical Panels.

MEMBER NEWCOMER: So we need to clarify again. This is not part of what we are voting on today, though. Is that correct?

DR. WINKLER: This is what you are voting on today. This is the composite measure. The weighting -- the criteria and the points, not the recognition part.

MEMBER NEWCOMER: I didn't see that in our documents.

CO-CHAIR FLEISHER: Bur the weights were endorsed by --

DR. WINKLER: They haven't been endorsed, not by NQF.

CO-CHAIR FLEISHER: No, by the

TAP.

DR. WINKLER: The TAP liked the weights, yes. when they reviewed them, they supported them, and suggested that this was a good composite of diabetes care, that the weightings made sense. They made clinical sense, and they supported the measure going forward.

MEMBER KEALEY: So what does the rejection of the less than 7 that we just did -- what does that do to this?

DR. BURSTIN: Nothing. So the NQF composite framework requires that all the measures within a composite be individually evaluated. They don't -- to be, rather, standalone or only as part of a composite.

So you have now indicated the A1c measure can only be as part of this composite or as NCQA agrees in that pairing other measure, but it is fine here, if you agree it is acceptable as part of a composite.

MEMBER KEALEY: And if we like

everything but the less than 7, is there any ability to parse that apart or --

DR. BURSTIN: You would have to again make those issues discussions back and forth with NCQA.

MEMBER YAWN: It is not a percent.

MEMBER AMARASINGHAM: But it would actually -- I assume, Helen, that that would actually mean we would turn down the measure, but suggest a change, because we couldn't approve the measure with the composite.

DR. BURSTIN: The measure being a composite.

CO-CHAIR FLEISHER: Vanita?

MEMBER PINDOLIA: I just had a question on the point system. How does that work? It is all or nothing? So let's say you have 16 percent greater than 9. You get zero points?

The way this is going to be used, I know in the state of Michigan and I know in other states, it is going to be used for the

physician incentive programs, and this is going to look so nice, because it is going to composite all of them together.

So all of a sudden, you have 16 percent. You get zero. So you are down to 90 right away. Is that -- or is there a grading? The 10 will become a 9 to 8, and the 7 to 6.

MR. HAMLIN: The total points add up to 100. If you can make 75 points, you achieve the SEQ ADA physician recognition for diabetes care, but that is -- This is the points, yes.

MEMBER PINDOLIA: But the points -- I am just wondering like for the 10, the 5, the 20 -- is it a 10 or a zero, a five and a zero? You either are there or you are not, or is it --

MR. HAMLIN: Yes, for each category. If you meet the criteria, then you get the points or not. Yes, sorry.

CO-CHAIR FLEISHER: Okay, David and Ruben.

MEMBER HOPKINS: So I have a point of order. Since NQF now has a requirement that we consider measures that are best in class, and since we have another type of composite measure that gets at many of the same things in a different way, I think it is important that we have the opportunity to review both before voting on either. I don't know how you want to do that.

DR. BURSTIN: Since there are different components, I think it would be helpful to finish our discussion of this measure and not necessarily vote, but then come back after the discussion of both composites.

CO-CHAIR FLEISHER: Yes.

MEMBER HOPKINS: Three different approaches.

MEMBER DELLINGER: If I am understanding correctly, the less than 7 here comes without the restrictions that were on the less than 7 that we just finished

discussing, because it doesn't say that anywhere here.

DR. BURSTIN: It says for special populations.

MEMBER DELLINGER: Oh, okay. Thank you.

MEMBER AMARASINGHAM: The question I have: This was clearly -- This was only developed by expert opinion. There is no empiric evidence to suggest these percentages. So one question I have is could this be proposed as a one-year time-limited, only because I am curious what is the underlying population in the United States where it actually achieved this.

MR. HAMLIN: This is actually -- This was actually developed through expert consensus but based on four years of data collection in the DRP, and this was just reviewed last year. It is based on four years of data collection.

MEMBER AMARASINGHAM: So do you

have a histogram or a distribution for how this would look, like if you got 80 percentile on this, what does that mean?

MR. HAMLIN: If you meet the -- Sorry, for the recognition? I am not understanding your question.

MEMBER AMARASINGHAM: Do you have a histogram of points?

MR. HAMLIN: Not with me, no, but they were just reviewed in 2009, and what they did is they looked at the data for each of these years and the weighting that came in from each of the provider offices that were seeking recognition. Then the experts basically judged that, yes, this weighting was still valid and usable and appropriate for this population.

CO-CHAIR DUBOW: This measure has been in use for a long time. It doesn't really qualify for time-limited endorsement. That is not what that process would do.

MEMBER NEWCOMER: Lee. I will

just make a comment. When we do come to voting, I would like to say that I like the measures and would strongly endorse them, but I would also amend, if it is possible, to get that criteria off the table.

What is important here is that you measure what is happening to a diabetic patient. You have a number of excellent measures there, and they are well thought out.

What kind of criteria table you might use to determine what is, quote, "good" diabetic care and other is nothing more than expert opinion, as already stated, and doesn't meet a good evidence standard.

These other measures, though, do meet evidence standards, and what is important is to report them.

CO-CHAIR FLEISHER: So when you say criteria table, do you mean the weighting?

MEMBER NEWCOMER: The point system that is there.

CO-CHAIR FLEISHER: The weighting

only?

MEMBER NEWCOMER: People could come up with 1,000 different point combinations, and they would only be opinion based. However, everything on that left tab is very good evidence based and are excellent measures to follow diabetic control.

MEMBER YAWN: Well, you are going to see another composite measure that takes those, and it just says yes or no. So you have to have 100 percent of all of them to get credit for anything, and I am going to tell you, it is a real bear to try to deal with 100 percent.

So I am not saying those percentages are right. I don't like only 60 percent of the eye, for example. That seems really low to me. But I think that at least they are trying to get at the concept of nobody is going to be perfect, and this at least gives you some --

MEMBER NEWCOMER: I think you also

made my point, that you don't think 60 percent is right.

CO-CHAIR FLEISHER: Helen would like to address the weighting system, and then we can go to the other measure. So let's let Helen talk.

DR. BURSTIN: Again, all these measures with the exception of A1c less than 7 have already been endorsed by NQF. So these are ones people can go ahead and report on right now. That is not an issue.

The only thing that is new here, in addition to the discussion we just had about A1c less than 7, is the idea of bringing in a composite. Again, in our definition a composite is combining multiple measures into a single score.

So by needing a single score, you've got to have some scheme that will bring them together. They could have -- whatever the case may be. They based on expert opinion this series of weights to get at what they

thought was most important.

Again, we could certainly go back to NCQA, ask for further details about the logic of it, but there is a requirement that we have to get to a single score. So unless it is an all or none composite, we've got to have some formula to get at that.

CO-CHAIR FLEISHER: A quick comment. Then I would like to look at the next measure.

MEMBER DELLINGER: A question: Are the percents in the center column there -- are those what have already been endorsed?

DR. BURSTIN: No. Those are thresholds that are not the issue. The issue is more -- So the actual clinical measures on the left have been endorsed, yes -- or right, whatever.

MEMBER AMARASINGHAM: I know that we have been saying that these measures have been endorsed, but it is not insignificant that there is new criteria.

DR. BURSTIN: No, that is the new measure.

MEMBER AMARASINGHAM: That is what I am saying. So I think --

DR. BURSTIN: That is just weighting.

MEMBER AMARASINGHAM: Right.

DR. BURSTIN: Essentially that creates the weights --

MEMBER AMARASINGHAM: But the weighting is very important.

DR. BURSTIN: The weighting is critical, and we want you to take a critical eye to it and see if it makes sense. Yes.

CO-CHAIR FLEISHER: So that is what we will or will not vote on, but let's go to the second measure.

DR. HALL: This is Bruce Hall. I am on the phone. I have a quick question. Is the weighting available? I am trying to follow the discussion and the materials, and I just cannot find this information.

MR. HAMLIN: It is not there.

DR. HALL: Okay. Thank you.

CO-CHAIR FLEISHER: Okay, and we haven't opened this up for public comment.

DR. HALL: This is Bruce Hall. I am not the developer -- as you may know, representing the college. But I am having a hard time following the discussion. How could any weighting scheme, whether it is an equal weighting scheme or any other weighting scheme -- how could that possible be submitted on the reliability of that composite if it hasn't already been in practice a long time, or the interpretability of those scores if they hadn't been in practice a long time?

CO-CHAIR FLEISHER: Thanks, Bruce, for that comment. We are going to move on to the second measure.

DR. WINKLER: The third diabetes measure is submitted by Minnesota Community Measurement. This is an optimal diabetes care measure. It is an all or none composite that

has five components, and we have it up there.

These are patients looking at the A1c level, the LDL level, the blood pressure, the tobacco use or nonuse, and the daily aspirin use.

So this is on a patient level data collection. How many patients have hit all five of the targets. So this includes patients 18 to 75 with diabetes who meet all of them. The Hemoglobin A1c is less than 8. The LDL is less than 100. The blood pressure is less than 130 over 80. They don't smoke, and -- this is the most recent revision -- for patients over the age of 41, daily aspirin use unless there are contraindications.

So it is a five-part measure that, at a patient level, if you hit all five, you get credit for it.

MEMBER JUSTER: Clarification. Wasn't this recently revised so it was a daily aspirin if you have cardiovascular disease or is it still at age 41 or above?

DR. WINKLER: Actually, we had -- Like I say, yesterday we got a --

CO-CHAIR DUBOW: It is. It is.

DR. WINKLER: Yes.

CO-CHAIR DUBOW: Thank you.

DR. WINKLER: Now but because they have made revisions that came in in the last day or two, the most recent one says daily aspirin for age 41-plus, use unless contraindicated. That is the most recent one that I got.

CO-CHAIR DUBOW: The material that we got --

DR. WINKLER: Right, and this came in like at four o'clock yesterday.

CO-CHAIR DUBOW: Okay. But it does make the distinction that Iver mentioned.

DR. WINKLER: Is someone from Minnesota on the phone? Excellent.

CO-CHAIR DUBOW: Because the Chairman of the TAP is not here, does somebody -- one of the staff want to just summarize

what the TAP -- their review, please?

DR. WINKLER: Yes. The TAP generally was concerned about a couple of aspects of it. When this was first presented, the revisions to the aspirin component had not been made, and there had been recent evidence to show the adjustment wasn't needed. So we had to wait for the changes that they have made.

It was felt that the A1c target of less than 8 was reasonable. The LDL was less than 100. Those are aligned with current re-endorsed NQF measures. So everything is fine. I think the biggest issue centered around the blood pressure target of 130 over 80. There were concerns, particularly, of most recent publications that actually were coming out the week the TAP met that this blood pressure target was of concern.

So that element of it was probably the major focus of the TAP's discussion and their concern with this measure.

MEMBER NEWCOMER: Reva, what was the concern? The data was showing it was too aggressive?

DR. WINKLER: Yes. Right. I think it was the most recent ACCORD trial, not showing benefit of aggressive blood pressure management for that population.

MEMBER HOPKINS: What is their response?

MEMBER NEWCOMER: Yes, have they modified it since then?

DR. WINKLER: Has Minnesota modified it? No.

MEMBER NEWCOMER: They have not. Okay.

MEMBER YAWN: I am also bothered by the 41 without evidence of cardiovascular disease. I do not believe that is evidence based, and for women it is 55, not 41. So I know that we have the one that says cardiovascular disease, but the update is the 41-plus.

MS. PITZEN: No, that is not correct. We just wanted to make a clarification as the measure developer, if we could.

Hi, this is Collette from Minnesota Community Measurement. We actually did work on this process, and the aspirin component is justification for cardiovascular disease, irregardless of age.

DR. JEWELL: Of what type of cardiovascular disease?

MS. PITZEN: We have a defined list of vascular disease, cardiovascular and peripheral vascular.

DR. JEWELL: Thank you.

MS. PITZEN: Thank you.

MEMBER NEWCOMER: So if the developer is on the line, are you intending to change the blood pressure recommendations or do they stay the same?

MS. PITZEN: I can answer that. We constantly are reviewing the evidence and

the guidelines. A year ago we revised our A1c target that was less than 7 to less than 8, and recently, based on the ADA standard and our expert guideline changes, we did change the aspirin component, and we would expect to be reviewing the blood pressure evidence as it emerges, and pulling our group together again to decide if we need to change that component. The evidence coming out is similar.

CO-CHAIR FLEISHER: So, Helen, can you comment? What happens if they change it after we endorse it?

DR. BURSTIN: If there is a significant or material change to the measure, it would come to NQF as an ad hoc review. We would together experts to review the change and make a determination if it made sense.

I mean, again, literally, the studies came out the day the TAP was meeting. So I think they were able to go back and do a revision on the aspirin one, but I think it is not clear they have been able to in literally

real time.

Let me just look up the date. The ACCORD trial was March 14th.

CO-CHAIR FLEISHER: Comments? B.J.?

MEMBER TURNER: So with moving targets on a lot of these variables, to have them all glommed together and have to meet them all seems like a pretty crude hammer to me. And I know that there is a lot of variability across racial groups.

I guess this is Minnesota, but I come from Philadelphia where we have a lot of African Americans where blood pressure control is much more difficult. You have to have four drugs sometimes to be at that level, and the side effects are much more significant, too.

There is a lot of concern about diastolic hypotension. We don't have any information about that here. So I think, within some of these, especially the hypertension and aspirin we are really

focusing on here, there are concerns and questions at this point.

I think, to have them all glommed together and you pass/fail is not a very sensitive measure, given that there is a lot of instability in the measure segments.

CO-CHAIR FLEISHER: Other comments? Questions for the developer?

MEMBER PINDOLIA: Just a question on the non-tobacco user piece. Can you just provide clarification. Is this purely the number of patients as a non-tobacco? How about if they are in a smoking cessation program?

MS. PITZEN: Collette. This is non-tobacco use.

MEMBER PINDOLIA: So it is just you have to be a non-tobacco user. So if you have someone who is currently smoking and is trying to quit, that doesn't count as a positive?

MS. PITZEN: That is correct.

CO-CHAIR FLEISHER: Comments? Lee?

MEMBER NEWCOMER: I will just offer the counter-view, that the thinking in Minnesota was that, since these patients -- if you get all five, you are going to get a much bigger bang for your buck in terms of delayed outcomes and complications. That is why they put the five together.

MEMBER YAWN: And the one they don't pay attention to is eyes.

MEMBER NEWCOMER: Right.

CO-CHAIR FLEISHER: Okay, are there any other questions regarding this particular measure, rather than comments on the appropriateness?

MEMBER AMARASINGHAM: I guess one question is can we endorse the measure with the requirement that it be revised to meet the core guidelines?

DR. BURSTIN: It could certainly be something where the measure could be

provisionally -- conditionally approved based on the potential for them revising the blood pressure measure, in particular. But again, that is one option for you.

DR. WINKLER: But realize that until they do, it is a No vote.

MEMBER ROSEN: I think, given the lack of attention to different risk groups, from what I have heard about the measure, is really very insensitive to the needs of the diabetic population, and it is really a very heterogeneous group rather than homogeneous. So I have a concern about that.

CO-CHAIR FLEISHER: Comment from the developer, or response?

MS. PITZEN: I don't have understand the last comment and what it meant.

MEMBER ROSEN: Well, you are just -- As Barbara Turner said, you are just kind of lumping in these five intermediate outcome or process measures and applying it to all patients with diabetes, when in fact patients

with diabetes have very different levels of risk and need.

So I don't think it is very sensitive to that.

MEMBER NEWCOMER: Although the evidence shows -- You know, we have had our debate about blood pressure, but the evidence shows that all diabetics would, in fact, have better outcomes with these measures. So I don't think it discriminates in that sense at all.

MS. PITZEN: Well, the other thing, as a developer I think we would like to just comment that, actually, as it is preventive, it does meet all evidence, and we wanted to be clear that Collette said earlier that, in fact, we have an eye to the new evidence preventive stuff. We review it. We will study the blood pressure in the studies, and it is among other studies, and I think it is important to improvement to review all evidence, not just base a change on one study,

and the 130 over 80 is supported in ADA guidelines.

MEMBER TURNER: I don't think we want to go into a long discussion about blood pressure cut points, but I do think that there is enough of a discussion going on in the world that you can't be dogmatic about the 130 over 80 is the measure we have to use for everyone.

CO-CHAIR FLEISHER: Thanks. I would actually ask Helen or Rita to give us some help in how we should proceed with talking about the two measures, as well as what else we need to do as far as voting for the individual criteria.

DR. BURSTIN: Sure. Just to make the point, as we were talking about earlier in terms of trying to -- You know, these are essentially competing composites. I don't know if one could imagine a world with both of them.

You would have to make that

determination, but what we would ask you to do is actually we want you to go through the formal process of voting on each of the criteria for each of them, and then make an assessment of what you would like to move forward, because we really do want to be able -- You know, the measures, in and of themselves, have to pass the criteria, and then you can make a determination of the next steps in terms of whether there is one that is best in class.

MEMBER YAWN: Could I ask the developer why the Minnesota measure does not include eye exams?

MS. PITZEN: At the time this measure was developed, and we have quite a bit of history back to 2003, it was felt that the intermediate outcomes of controlling enough the LDL and blood pressure would prevent potential complications down the road from occurring. So we did not include that process measure as part of our composite.

MEMBER YAWN: My comment would be that is nice down the road maybe 10 years, but it may not be nice for people between the time you start and 10 years down the road, because they already have eye disease.

CO-CHAIR FLEISHER: Iver, and then I would like to move to taking the vote. Last comment.

MEMBER JUSTER: And, actually, I suppose the same comment could be made for nephropathy screening.

CO-CHAIR FLEISHER: Okay. So we are going to actually discuss the measure we just discussed. So we will start with the Minnesota measure, and Reva will take us through voting on each of the criteria.

DR. WINKLER: Just as we did with the other measures in the first group, we do need to have the Committee's final assessment on how well the measure meets the criteria for importance to measure and report, scientific acceptability of the measure properties,

usability and feasibility.

I will draw your attention to the handy dandy little cheat sheet provided at each place of those criteria. But having this sort of data helps as the measure moves through the process to kind of provide summary evidence for the various audiences going forward to help determine how things are going to go.

CO-CHAIR FLEISHER: Can we put the TAP's recommendations for these?

DR. WINKLER: The TAPs only looked at the sub-criteria. They did not vote on the overall criteria. That is your -- That is a Steering Committee role. But how they felt on the sub-criteria are listed here as well.

All right. So we are looking right now at the Minnesota Community Measurement composite measure, optimal diabetes care. So what we are looking for is your -- the Committee's assessment on importance to measure and report. This is a

yes/no vote.

So do you want to do it? Do you want me to do it?

CO-CHAIR FLEISHER: All those that vote Yes?

DR. WINKLER: Eighteen.

CO-CHAIR DUBOW: Noes?

DR. WINKLER: Three.

CO-CHAIR FLEISHER: Abstain? One.

DR. BURSTIN: And people could actually refer to that little handout. It would be helpful, because again we are asking you to do an overall assessment of importance based, in fact, on the three sub-criteria of the impact to the condition, if there is a known gap in care and the relation -- and the third piece of that would be evidence, and I guess that would be potentially part of what people are voting on. So to be clear.

DR. WINKLER: So the next criterion is scientific acceptability of this measure as it is specified, the properties of

this particular measure, and there are multiple sub-criteria around the specifications, reliability, validity, their exclusions, risk adjustment, meaningful differences, comparability and approach to disparity.

You can see the recommendations of the TAP in terms of those sub-criteria. So the Steering Committee will provide the overall rating for that criterion.

MEMBER YAWN: Sorry. Can I ask just what about -- In scientific acceptability, I don't see a place for gaps.

DR. BURSTIN: That is under importance to measure and report. You have already voted on that.

MEMBER YAWN: Thank you.

DR. BURSTIN: You are welcome. This is really the scientific acceptability of the measurement properties, its reliability, its validity, the precision of the specifications. That is not the evidence.

You just did that one.

MEMBER HOPKINS: So do we have the data on reliability and validity or did the TAP -- The TAP looked at it and said it was complete?

DR. BURSTIN: Yes.

CO-CHAIR FLEISHER: So those who vote Yes for this criterion? No?

CO-CHAIR DUBOW: No.

DR. WINKLER: Give me a second. Importance actually is a threshold criteria. You voted Yes. So we can move forward. I you had voted No, we would stop.

Okay. So on scientific acceptability, the voting is: Completely adheres to the criteria; partially adheres to the criteria; minimally adheres; or not at all. Okay?

How many would favor completely meets the criteria for this measure? I am seeing one.

MEMBER NEWCOMER: Clarify again

what we are voting on.

DR. WINKLER: This measure as specified completely meets the criteria as listed and laid out by NQF for scientific acceptability of the measure properties, this measure as specified. Would that be partially?

So we try it again. How about completely? Anybody to vote for completely?

How about partially? All right. Twenty-one. Okay.

How about minimally? I knew there was at least one. Okay. And abstentions? I don't believe so. I think we've got everybody. All right.

The next topic is usability of this measure. The usability criteria focused around distinctive or added value, how well it is harmonized with other measures, and provides added value as a new measure to NQF's portfolio. Okay?

How many think it adheres to the

criteria completely? Six?

Partially? Thirteen.

How about minimally? Two.

Abstain? No.

Okay. The last criterion is feasibility, and this is what is the burden, what is the cost, what does it take to generate the information, particularly around how well it is amenable to use by electronic sources, moving into EHRs, whether exclusions require different data sources, potential for inaccuracies or errors.

MEMBER NEWCOMER: Are the comments up there relative to feasibility or are they more relative to usability?

DR. WINKLER: Right. Well, sometimes we would get -- It gets to be a messy discussion sometimes.

So the vote for feasibility. How many think it meets the criteria completely? Fifteen.

Partially? Five.

Minimally? One, two. Two.

Abstentions? Zero. Okay.

CO-CHAIR FLEISHER: Do you want to do the other one?

DR. WINKLER: Yes. I think it would be confusing to do the overall? I think you need your recommendation.

DR. BURSTIN: This is a little unusual. I think the simplest thing is let's rate the criteria. Then you could see the criteria as you have rated them head to head, and then make your assessment, I think, would make the most sense.

So let's do the other one, same game. Let me explain that, actually, because the logic is we would only -- I'm sorry.

I think the issue is we would only do the head to head comparison if, in fact, both meet our criteria. So we need to establish that first, and then we will come back and do the assessment.

DR. WINKLER: Now we are going to

turn our attention -- Hawa, do you have the table up? Great -- to the NCQA composite. If you recall, it had the numerous process and outcome measures, I believe, on that page.

All right, so important to measure and report. This is a yes/no criterion. So how many would say, Yes, it is important to measure and report? A unanimous 22.

All right. Scientific acceptability of the measure properties as specified, and that includes the composite methodology of the weighting and the points. All right?

So how many feel it meets the criterion for scientific acceptability completely? Okay.

Partially? That is 17.

Minimally? Four.

Anybody else? Not at all? No? Okay.

In terms of usability of this measure -- What?

MS. BOSSLEY: I think it is five for partial, because we should have 22 voting. I meant minimal. I'm sorry.

DR. WINKLER: No, there is five -- I was going to say, there were six, not even five. All right. Given we did so well with that one, let's try usability.

How many would say it meets the usability criteria completely? Four.

Partially? Fifteen.

And minimally? One, two, three.

MEMBER PINDOLIA: Can I change my vote to a higher one?

DR. WINKLER: You were partial?

DR. BURSTIN: So you want to go to Completely?

MEMBER PINDOLIA: Yes.

DR. WINKLER: Okay. So you are 14. We can do that. Five-fourteen-three. All right.

Feasibility criteria: Completely? Eleven.

Partially? Eleven.

Minimally? Is there anybody still on minimally? No? It should be everybody. Right? Okay. All righty.

CO-CHAIR FLEISHER: Okay. Helen?

CO-CHAIR DUBOW: Heidi is doing it.

CO-CHAIR FLEISHER: So next do you suggest voting on each individually, Helen?

CO-CHAIR DUBOW: If you give me just a second, I will give you a summary of what you said.

MS. PITZEN: Hello?

DR. BURSTIN: Sorry. We are just compiling results. You haven't missed anything. Just give us a moment.

DR. WINKLER: Okay. I have -- I can tell you. So let me start first with the all or none. It is not pretty.

CO-CHAIR FLEISHER: While they are putting this up, I just would propose -- Does anybody feel these are competing versus non-

competing measures? In particular, does anyone feel that they both could be endorsed, because they are non-competing? David, do you want to make a comment?

MEMBER HOPKINS: I feel they are competing, and I feel that there is a property of the Minnesota measure that really hasn't been highlighted. It is, in fact, a patient-centered perspective measurement, and we have so few measures that are formulated this way. I just think it is really important for people to think about that.

It is about the whole patient getting the care and with the results that are important to that individual. So I strongly favor that measure.

CO-CHAIR FLEISHER: Iver?

MEMBER HOPKINS: I think the flip side of that is I do find that in the other measure the need for weighting invites arbitrariness, and there just is no way to scientifically validate the weights.

MEMBER JUSTER: There is a compelling case to be made for all or nothing measures or 80 percent or more measures or whatever you want to call them. I don't know if there is space to say something about -- I might be happier voting yes if this measure was harmonized or it included the retinal exam and screen for nephropathy.

So it is sort of a great idea "and" kind of thing. Is there space for that sort of thing?

CO-CHAIR FLEISHER: Patchen?

MEMBER DELLINGER: Just in terms of how we describe things, we are calling the Minnesota measure an all or none. It is all or none for individual patient, but the measure is the percent of patients who meet the all or none criteria.

CO-CHAIR FLEISHER: Brian?

MEMBER FILLIPO: I also am in strong support of all or none criteria. I think there are patient centered, although I

think this specific collection of indicators is just not granular enough, and not risk adjusted. So I don't think that this particular one is good.

CO-CHAIR FLEISHER: Do you make a comment, Joyce?

CO-CHAIR DUBOW: I was just repeating something I heard at the table and that is whether, in fact, these are two measures, because they set different levels of achievement, the all or none measure as opposed to one that is one of gradation and so whether, in fact, these represent two measures.

On the other hand, if the idea is to represent to the public, for example, to a patient what is good diabetes care, then we have a different cut. But the issue is whether we have the opportunity to make room for both or whether we need to take a cut on it.

CO-CHAIR FLEISHER: Patricia?

MS. HAUGEN: Yes. I just wanted to comment. If you look at usability, it is that is this understandable to the public, and can it be used in decision making? I think that speaks to the Minnesota measurement where, although it is the percentage that there is some clarity in this, I think one issue of composite measures from a patient perspective is how you really understand the intricacies of it, and can I make a decision or does it inform me? I think this one has some clarity to it that, from a patient standpoint, would make it usable.

CO-CHAIR FLEISHER: Barbara?

MEMBER YAWN: But I am going to go back to you say it is patient-centered. It is patient-centered as long as you don't care about their eyes or their kidneys. But there's -- Well, outcomes are identifying early eye disease, and being able to prevent blindness. I think that is very significant. So I am concerned about those lack. I

understand the simplicity.

CO-CHAIR FLEISHER: So I would actually ask Reva, for example, would we ever endorse a measure but say we believe that there is -- from either a research or a gap in care perspective, these two are not -- have been ignored, and it should be included in another measure to be developed over the next year? Would that be reasonable?

DR. BURSTIN: There is only so much you can do in terms of conditions on a measure. The measure that is before you, not conditions, but that would be essentially rejecting and asking a different measure to come back.

You can potentially ask for harmonization of the individual components, if that is something you would feel more comfortable with. That is something, I think, within your purview, but I think adding components to the composite couldn't be done at this point.

I mean, for example, this issue of the blood pressure control -- There are, you know, the aspirin measure. I think there are potentially conditions that could be placed on it, if you feel it is important, and that is, in fact, we do -- although I am not sure we should, we do specifically have criteria for composite measures, and harmonization of the components is an important aspect. So just another consideration.

MEMBER AMARASINGHAM: Can I ask a quick question? When we vote on the all or none measure then, will there be an option about the condition with respect to the ACCORD trial guidelines in harmonizing the blood pressure?

DR. BURSTIN: Yes.

CO-CHAIR FLEISHER: So we will include that in the vote, conditional pending blood pressure, as one of the options. Then secondly -- Okay, do you want to propose what the condition would be, so that when we vote,

we know what the condition is?

MEMBER AMARASINGHAM: I am not familiar with the specifications of the ACCORD -- you know, what the results were from the ACCORD trial with respect to this. So I guess I would say that they should reflect the results of the ACCORD trial, which I believe --

CO-CHAIR FLEISHER: It is the interpretation.

MEMBER AMARASINGHAM: Right, but at least a review of it.

DR. BURSTIN: Right, and I think it would be reasonable for you potentially ask that they respond back, for example, why the blood pressure -- I mean, I think their response we have heard on the telephone was that it was one study. They had looked at the full view of the evidence.

I think the concern is, as I recall from the TAP, and I did just pull up the study, is I think the issue was the fact

that there is possible harms associated with a lower blood pressure, and that was the issue they really homed in on. It wasn't just an issue of the fact that it didn't significantly reduce the bad outcomes from cardiovascular disease, but that there were signals of possible harm with intensive blood pressure control, including a rate of serious adverse events that were significantly higher in the intensive therapy group compared to the standard therapy group. That was the issue they are really homing in on, was the safety issue.

CO-CHAIR FLEISHER: So can we actually say conditional and then reviewing and responding to the TAP's concerns, and that would be what I would propose as the condition, so that they could either change it -- approve it by either changing the blood pressure or responding appropriately, and there's multiple other levels in which this needs to be approved. Okay?

MEMBER PINDOLIA: I'm sorry. I have been trying to read the results. I was trying to figure out exactly how many people are meeting these five, and out of their 119,000 patients submitted, 19 percent met all five targets with a range of 45 percent to below one percent.

So does that mean below one percent physicians are considered to be not practicing properly or is it they just have a negatively selected patient population or in an urban area or --

MEMBER NEWCOMER: It simply means they had a 99 percent opportunity for improvement. That is all it means.

MEMBER YAWN: Not when they get paid.

MEMBER PINDOLIA: Right. Not when there is going to be incentives applied to it and not when there is going to be -- just because you have higher smokers, because you don't have an opportunity to even have smoking

cessation.

DR. BURSTIN: The measure is titled optimal care, just to recall that. It is inherently a different construction. I just want to be clear. That is actually how it is titled.

MEMBER HERMAN: And the developers are on the phone, and they could tell you that, when they have looked at it, it is not stratified by one group or another. I mean, there are some clinics that should really be doing a whole lot better, but are doing very, very poorly. So there is not a bias within this that we can detect.

MEMBER PINDOLIA: That could be used like to then have -- like how is that going to be used? It is just a reporting or there going to be actual them saying what does your practice need or what is it missing? Is there processes to help improve those?

CO-CHAIR FLEISHER: So I am going to call the vote.

MEMBER NEWCOMER: There has to be a comment here. If we are still looking between the two measures -- We are not then? Okay. All right.

My comment: The difference between the two measures --

CO-CHAIR FLEISHER: Do you want to go through the criteria, Reva?

DR. WINKLER: On the optimum care, the ONM measure, you said it was important. It partially meets the scientific acceptability criteria, partially meets the usability criteria, and completely meets the feasibility criteria.

Compared to the NCQA composite of the multiple measures: Yes, it is important, partially meets the criteria presented for acceptability, partially meets the criteria for usability, and partially meets the criteria for feasibility.

MEMBER YAWN: So the only variant is really the feasibility.

CO-CHAIR FLEISHER: And the scientific acceptability.

DR. BURSTIN: Can you go back?

MEMBER NEWCOMER: There is a key difference in philosophy between these two measures. The first measure, the five aggregates strives toward as best performance as you can possibly get. Ultimately, your goal is to get as close to 100 as you can.

The other is about minimal performance. It is more like meeting test standards. If you get to your 15 percent, you can move on to the next measure and improve it, my point being, if you take a look from a clinical standpoint, I think measure one says we are going to try harder to get to total perfection.

The second measure is more about passing the test, and for that reason I would favor the first measure.

CO-CHAIR FLEISHER: But realize, you just actually advocated that you could

have both..

MEMBER NEWCOMER: You certainly could.

CO-CHAIR FLEISHER: I mean you could actually endorse both measures of what I am hearing, and actually say did you pass the test, and did you have a perfect score, which are two different criteria.

DR. BURSTIN: Just one other point. Since they obviously are so close in terms of your ratings, there is really not a whole lot of light between them as you actually look at the ratings of them.

You are still fairly early in this process. So, again, there is also nothing wrong with potentially putting them both out for comment and allowing us to get what is usually a very -- I assume we will get a fairly robust response on this, the set of two measures, is one potential idea.

CO-CHAIR DUBOW: I like that idea a lot. I also think that it is possible that

these two measures appeal to two different interests: One, the patient who, obviously, wants optimal care and is going to be very, very interested. It is intuitively easier to understand whether you have got everything you should get, and in only one measure. The Minnesota Community measure absolutely does that.

On the other hand, it seems to me that the NCQA measure does have some value from a clinical perspective, ticking off the items, because you are knowing whether you are getting past some predetermined threshold.

So it feels to me as though there is the potential here for having our cake and eating it, too, in that there is one measure that really will resonate with patients, and a second one that may have some more salience for the clinical audience.

CO-CHAIR FLEISHER: And that may get back to the question of what is your population in the clinic. If the population

is highly motivated, then you may be able to achieve optimal care. If the population is much less motivated, then a measure that says you are doing certain things and can you actually pass a test, given a less motivated population? may be important.

Last comment, and then we are going to vote.

MEMBER HOPKINS: So Joyce's comment reminds me of our recent debate about the -- and the Board's debate about NQF expectations for use of measures, and I believe where it sits is public reporting -- what we call accountability and quality improvement, and the important word there is "and."

So I personally favor the first measure for public reporting, and I do not favor the second, because it is too low a bar.

CO-CHAIR FLEISHER: It requires the patient to be involved.

MEMBER YAWN: I am still going to

argue that optimal care doesn't mean the process measures for some outcomes are ignored while the process measures for others, like Hemoglobin A1c which is a process -- it is an intermediate outcome, just like having an eye exam is an intermediate outcome. It is not optimal care, in my opinion, but you have all said you think it is.

CO-CHAIR FLEISHER: So we are actually going to vote, I think, yes or no on Minnesota, followed by yes without conditions, yes with conditions that they respond to the TAP regarding the blood pressure issue, no, and then a second comment with regard to potential gap in care unless we just rather we would endorse that they evaluate in the future including issues of retinopathy and nephropathy. Would that be fair? Okay.

MEMBER HOPKINS: And the Minnesota one has the condition that they look at the blood pressures.

CO-CHAIR FLEISHER: That is a

minimum. That is the second. So the first one is the NCQA, which is a simply yes or no. I'm sorry. I made the mistake in the order. Yes, my fault.

CO-CHAIR DUBOW: Oh, okay.

CO-CHAIR FLEISHER: Okay. I was looking at that and remembered. So we discussed NCQA first. All those in favor of endorsing the NCQA measure?

MEMBER NEWCOMER: We can do more than one?

CO-CHAIR FLEISHER: Yes.

CO-CHAIR DUBOW: We can. We can do more than one. Yes, we can. Sure.

DR. BURSTIN: We can do more than one, because at this point all you are doing is approving it to move for the membership.

CO-CHAIR DUBOW: What was that?

DR. WINKLER: I got 13.

CO-CHAIR FLEISHER: This NCQA is not with conditions.

DR. WINKLER: Okay. Thirteen,

okay.

CO-CHAIR FLEISHER: All those who vote No?

DR. WINKLER: Eight.

CO-CHAIR FLEISHER: Okay.

DR. WINKLER: Are there any abstentions?

MEMBER ROSEN: I just wondered if we could, for the NCQA one, have a condition that we get some sort of more empirical evidence as to the way they system has been used.

CO-CHAIR FLEISHER: Is that a condition or just a comment, because I didn't hear any -- We voted Yes without conditions. It is a comment.

MEMBER ROSEN: It is a comment.

CO-CHAIR FLEISHER: Okay. We need to vote. For those who voted Yes, is it vote Yes without conditions -- well, that is what I heard. No?

DR. WINKLER: I hadn't heard the

conditions. The question, I think, Amy just brought up was if -- your comment, you would like to see more data, but does that mean that, if they don't produce it, then you vote no?

MEMBER ROSEN: I am concerned about the weighting and how the criteria were developed for the weighting. So if they can produce more evidence on that, that --

MEMBER NEWCOMER: We have already been told there isn't any. So they aren't going to be able to do it.

MEMBER ROSEN: It is just clinical judgment that something is a 20 versus a 10. Expert opinion, and in my mind that is not good enough. So, okay.

CO-CHAIR FLEISHER: Well, you voted Yes or No?

MEMBER ROSEN: I voted No.

CO-CHAIR FLEISHER: You voted No. Okay. So then we can just give that comment.

Okay. Next is Minnesota. So now

-- Oh, are there any abstentions?

DR. WINKLER: There would have to be. We need 22.

DR. BURSTIN: Were there any abstentions on that vote?

DR. WINKLER: Oh, Barbara is out.

CO-CHAIR FLEISHER: So we will get it when she comes back.

Okay. Minnesota: So we have Yes without any conditions. Remember there is the issue of the ACCORD trial. So anybody vote Yes without any conditions?

DR. WINKLER: One, two. Two.

CO-CHAIR FLEISHER: Yes with the conditions that they evaluate the ACCORD trial and respond regarding the appropriateness of their blood pressure criteria?

DR. WINKLER: Fourteen.

CO-CHAIR FLEISHER: No?

DR. WINKLER: Four. Any abstentions? No? And Vanita. One.

CO-CHAIR FLEISHER: Okay.

MEMBER YAWN: And also Barbara. She is not here.

CO-CHAIR FLEISHER: And a simple -- Do we need a Yes or No that we would like them to look at nephropathy or retinopathy or are people willing -- So we don't vote on that. We just suggest?

Is it the will of the committee that they look at that for any future measure? Okay. Thank you.

DR. BURSTIN: Could you go over the final votes, Reva?

DR. WINKLER: The final votes were Yes for the measure as is; with two Yes with the condition that they respond to the blood pressure, the ACCORD trial, and come back to this committee. That was 14. Noes were four.

I think what is going to happen is, when they come back, we will probably have to re-vote it.

I think there is time for us to get that back so that it isn't left unresolved

and hanging. I think the better we can get it resolved, it would be useful.

CO-CHAIR DUBOW: All right. I think we are done with the diabetes measures. Is that right?

Okay, the first one is always harder. That was very difficult. Now we have to make a decision. I can't imagine that some people don't want to take a five-minute break, but it is also close to lunch. So maybe we can combine -- Okay, we are checking on lunch. I was going to suggest that we take a quick break, bring our food back and -- It is not there? Okay. So it will be at one. Well, can everybody manage 20 minutes? No?

CO-CHAIR FLEISHER: So why don't we just have Reva actually go over the measure?

CO-CHAIR DUBOW: Yes. Well, what we have been asked to do is to -- We are back to talk about the cross-cutting measures, and we were going to hit the BTE measure first, if

that is okay. But if you want to introduce all of the cross-cutting measures, Reva, that is what we should try for. But we will start with that one.

Okay. So we are about to start our discussion on the cross-cutting measures. We will stop at one, and next time build in a little bit of a break.

DR. WINKLER: So we are going to do this one first. The first one we will be looking at is a measure that is the proportion of patients with a chronic condition, and they are listed in there, multiple chronic conditions, that have a potentially avoidable complication during an entire calendar year.

I think it is important -- I hope you have had a chance to kind of read the Word document that came along with the measures talking about the background of the development of these measures as part of the Prometheus Project.

The potentially avoidable

complications is a concept that these are being built around. There are numerous measures -- we are going to see three others -- around more shorter term acute events, but this is the cross-cutting for chronic conditions, looking at avoidable complications.

Some of them include readmission. Some of them include types of care required for things like DVTs and sort of other kind of complication type measures, as well as utilization type things such as readmission or ED visit or all of these sorts of things.

So that is the measure, and I guess having read that background piece will help better understand the Bridges to Excellence folks' approach and, as I said, this is all part of the Prometheus payment system. So that is the first one.

The second one is the Medicare Health Outcomes Survey, the physical components summary score. This is -- The

measure steward is NCQA. This is a patient survey tool. It is in use in the Medicare population. It is based on the SF-36 -- I think the A version -- but this is a measure that has been around a long time.

What we gave you, in addition to the submission were the links to their very extensive website that has the tool and the current results and a lot of the research that has been done around it, those sorts of supporting documents.

The third one is another measure from NCQA, the care for older adults. This is a three-part measure that includes results for advanced care planning, functional status assessment, and pain screening. But these are three parts with each one reported as an individual measure and with individual results.

So those are the three cost-cutting measures we are going to be looking at on the next go-round.

CO-CHAIR DUBOW: Okay. We have -- We are going to do the BTE measure first. So, Francois, will you join us at the table? We have the developer with us, and if you would like to make some opening remarks, that would be just fine.

MEMBER NEWCOMER: This is the 2209 measure?

CO-CHAIR DUBOW: Sorry. This is the -- Right.

MR. DeBRANTES: Well, yes, thank you. I don't know if, Reva, there was an opportunity to send out the PowerPoint slides.

DR. BURSTIN: It is on the PDF.

CO-CHAIR DUBOW: So let's just find it. Do we have it? Just bear with us for a minute. No?

CO-CHAIR FLEISHER: I don't see it.

CO-CHAIR DUBOW: Well, she will pull it up.

MR. DeBRANTES: All right. Thank

you, and I apologize for those not having been sent, I guess, on time.

What I wanted to do is to summarize a little bit what our approach has been in devising this comprehensive complication of care measure and, obviously, be in a position to answer questions that any members of the Steering Committee might have.

As Helen mentioned, there are others like this coming up a little later. So they are all constructed essentially the same way, although this one measure is around chronic conditions, in particular. The other ones are around acute medical conditions.

We have been through four TAPs. I would say that the grade we have gotten from the four TAPs are highly variable as, by the way, are potentially avoidable complication rates. So I think it is well consistent and probably speaks to, I think, both the definitions around the measures and the challenge that they represent for most

physicians.

Just a couple of words on where this all came from which, as Helen mentioned, is a by-product of our work around creating episodic care payment, and in doing so, part of our charge has been to determine the extent to which we can identify care that is appropriate, right for patients, versus things that might happen in a patient's life that are caused by what we have come to call potentially avoidable complications.

What is important in these definitions and in this measure is that we think about this as truly a patient centric measurement. So this isn't about an individual physician or an individual hospital. This is about a patient.

In some instances, that patient might have asthma but with comorbid conditions, or that patient might have diabetes with comorbid conditions.

We don't try to parse out the

diabetes from the asthma, from the COPD, more than to think about the patient as a whole, and to determine the extent to which the patient has had one or more avoidable complications during the course of their episode, which we define in the case of chronic conditions as being one year.

So I think one of the -- bringing back some of the comments that we got from the first TAP we went through, which I think was pulmonology, there was a significant pushback that the measure wasn't tightly linked to COPD or asthma. That is because we don't think about this from a provider centric perspective.

We think about this from a patient centric perspective, and a patient with COPD often has comorbid conditions. So the extent to which they are hospitalized for one of their comorbid conditions as opposed to their core condition, we consider that a potentially avoidable complication, because what we are

trying to do is to create a measure that looks at and evaluates the system of care around the patient and creates accountability for the system, not necessarily individuals but anyone within the system, and creates co-responsibilities for all the providers that manage and co-manage, whether they do it consciously or unconsciously, or don't do it, but are supposed to be co-managing the patients.

So the potentially avoidable complications are essentially divided into three parts, and maybe we will skip to that section so that folks can understand it.

Well, let me just quickly go through this. So we have gotten lots of help, actually -- and this is just a few examples; there is a larger summary in the Word document that was sent out.

We have had a fair amount of support from AHRQ, CMS, HC Health Partners, other organizations, ACC-related physicians,

in really doing and looking at this issue of what can you consider to be typical care versus care associated to potentially avoidable complication, and all this help from physicians across the country has been baked into these definitions. Next slide.

So the six chronic conditions that we have studied are the six that are listed here: Diabetes, coronary artery disease, congestive heart failure, COPD, diabetes, and hypertension.

Importantly, this measure, as you have looked at the definition, is really for patients below the age of 65. So I want to make sure that that is clear. We are not including patients above the age of 65, which often happen to have multiple, multiple chronic conditions, mostly because we haven't studied patients above the age of 65.

We study patients below the age of 65, and so that is what this measure is intended to look at, is accountability in the

management of patients under the age of 65 in commercial populations. Next slide.

So here is the way an episode of care is defined. The reason why I bring this up is because we count avoidable complications during the period of time that is defined around this episode. In this instance, it happens to be one year.

So a chronic care episode is looked at as being a one-year episode. We look at it usually in a calendrical fashion so that it is tied to the benefit year. It is tied to contracts. It is tied to a whole bunch of other things.

The claims that are analyzed as part of any effort around measurement using claims are really distinguished between two types of claims, professional services, labs and other ancillary services and drugs -- outpatient based claims, if you will -- versus inpatient based claims.

So these come in two different

streams, and we accumulate them together. As we look at an episode of care during the course of a year, we classify events, services, as either being typical -- so those are the ones that are illustrated on this chart as blue -- versus potentially avoidable, and we will get into the definitions of what we consider to be potentially avoidable.

I emphasize potentially, because we don't pretend that any of these are completely avoidable, but potentially avoidable and should, therefore, be worked on.

Inpatient stays and ED visits, in our definitions, are mostly potentially avoidable complications for these patients with chronic conditions. Next slide.

So the way you arrive at the measure and the measure definitions and the accounting around the measure is really using claims data. So all claims data come into -- this is just an illustration of a funnel. Some get excluded. Why? Because they are not

at all relevant to that patient's condition, and cancer is an example. Then the balance get sorted between typical services versus potentially avoidable complications or services associated to potentially avoidable complications.

The measure that we are proposing here, and with the three that will follow, is counting these events, so counting events that are associated to avoidable complications. Next slide.

So just to give you a sense of the size of the database, the sample sizes of the patient cohorts that we have analyzed through our effort, very large amounts of patients. So four million in total, 172,000 patients with diabetes. We do not suffer in our analysis from small sample size problems. We have very adequate sample sizes to do severity adjustments and other analyses.

MEMBER YAWN: Lower age limit?

MR. DeBRANTES: Excuse me?

MEMBER YAWN: Lower age limit?

MR. DeBRANTES: Lower age limit depends on the condition. For the most part, it is 18, except for pediatric asthma.

MEMBER YAWN: Well, I assume that will go down to two or three.

MR. DeBRANTES: So in a snapshot, you can see that, out of the total 650 or so patients with these chronic conditions, about 72 percent during the course of a year had one or more potentially avoidable complications. Again, these can be associated to either the core condition, comorbid conditions or patient safety issues, and I will get into that.

There is a huge amount of regional variation, as you can imagine, for rates of potentially avoidable complications by condition. This is just a snapshot. We ranked all of the states into deciles, and then we ranked the states by decile, so that you have a decile distribution here with the min, the max, and kind of the average.

I think what is important is that -- and we see this not just in these chronic conditions but also in the acute medical hospitalizations -- is that there is more distribution, I would say -- or there is a wider variation on the top end of the decile distribution than there is on the lower end, again not surprising, but it does tell us, I think, clearly, in our analyses that a significant percentage of these avoidable complications can, in fact, be avoided; because you have -- The variation is not explainable by severity of patients.

So adjusting for the severity of patients, you have these significant variations in rates of hospitalizations and emergency department visits by patients.

When we actually look at the distribution of these patients across the country, unsurprisingly and related to your prior discussion, patients in Minnesota on average have far fewer rates of potentially

avoidable complications than patients in, say, Arkansas, and I am not passing judgment on one versus the other more than I do think that there is a relationship between low rates of potentially avoidable complications and good systems of care. Next slide.

So rates of potentially avoidable complications fall into three categories. Type 1 are avoidable complications that are associated to the patient's core condition; so diabetes, for example.

The second would be related to comorbid conditions. So a patient with diabetes also has asthma. If they are hospitalized for their asthma, we will consider that a potentially avoidable complication.

The third one are avoidable complications that are associated to patient safety issues, and an adverse drug event is a classic example of such a potentially avoidable complication.

When we look at the distribution of those types of avoidable complications by disease category, the vast majority of avoidable complications come from Type 2, which is comorbidities and patient safety failure. Far fewer come from the core condition.

I think this speaks to the critical importance of this measure and its contribution to the field of accountability, because what we see is that, when you are looking at the tight scope of a measure, asthma or COPD, and you are trying to determine the extent to which that asthma or that COPD is being controlled appropriately, we find that on average, yes, although there are avoidable complications. But what seems to be forgotten are the comorbid conditions of the patient, and that is what is driving a lot of these avoidable complications.

So if we make this move from a solid measurement of individual physicians to

as more systems measurement of surrounding the patient and understanding whether or not the system is, in fact, working to the benefit of the patient, we have to look far broader than just the tight Type 1 potentially avoidable complication.

Some further scopes again -- and these now start relating back to dollars, because we do look at dollars quite extensively in our work. So that both the volumes or the frequency of avoidable complications, but also the costs associated to these avoidable complications are highly skewed toward PACs of Type 2, so associated to comorbid conditions with 255 out of $400 million. So lots and lots of dollars being spent in this population on avoidable complications. Next slide.

This is just somewhat of a top 10, if you will or close, of top drivers of potentially avoidable complications for each one of these chronic conditions.

Emergency room visits, as you can see, are a significant driver, but then again the acute flare-ups of the core condition, UTIs, cardiac dysrhythmias, pneumonia, lung complications.

So these are the things that most of you who practice know that, when patients hit the hospital at the emergency department, this is what they present with. This is what we see as avoidable complications.

The reality is that, if the patients are fundamentally managed as a whole, I do think that we can legitimately say these numbers should go down, and without holding the system accountable for them, then it ain't going to happen.

We do have relatively robust severity adjustment built into each one of these measures, and this is just an illustration of how that severity adjustment works.

So if you've got the population

severity index that varies, that is going to have an impact on their rate of potentially avoidable complications, and you can recalculate the rates of potentially avoidable complications severity adjusted to the patient population, so that when you are using this for comparative performance purposes, it is fully severity adjusted.

MEMBER TURNER: Which measure is it?

MR. DeBRANTES: Excuse me?

MEMBER TURNER: Which measure are you using -- the severity level?

MR. DeBRANTES: The severity adjustment is based on a -- Excuse me? Well, we used some of it, but it is really a relatively standardized linear multi-variable linear regression model.

DR. RASTOGI: So we created the risk adjusted models for each of the six conditions, and then calculated for every patient their severity score, and then for the

whole population you can calculate a severity index.

So then you compare one health plan to the other health plan or one employee base to the other employee base, whatever is their population. Their constituent base, we will call it as one, and then each other population we can look at the sum total of the severity scores of individual patients with that condition.

So it is all based on a linear regression model like Francois was mentioning.

MEMBER TURNER: So I am just trying to figure out what the inputs are into the severity model.

MR. DeBRANTES: Well, resource use and cost of care as an initial input.

DR. RASTOGI: Yes. So diagnosis scores, pharmacy goes into it. Procedures go into it. Quite a bit of the services that are there, they all go into the severity adjustment, and it is all part of the measure

development.

We have included all the variable lists in the measure submission.

CO-CHAIR FLEISHER: Amita, would you introduce yourself, please?

DR. RASTOGI: Yes. I am Dr. Amita Rastogi. I am Medical Director with BTE, work with Francois.

MEMBER HOPKINS: Can somebody explain why the risk adjustment didn't work inversely on this slide? If coefficient B is higher severity -- am I interpreting that right? -- then why wouldn't you downward adjust their PAC rate?

DR. RASTOGI: If it is higher severity?

MR. DeBRANTES: No, it is the opposite. It is lower severity. So the index work in the opposite way, but yes.

DR. RASTOGI: No, that is exactly right.

MR. DeBRANTES: This is actual

calculated severity index.

MEMBER HOPKINS: So lower index is higher severity.

MR. DeBRANTES: Right. You can do it either way.

MEMBER TURNER: Has this been published anywhere?

MR. DeBRANTES: Yes. Actually, we have published it in a couple of journals, and we have one new paper that is in review now with HCR.

CO-CHAIR FLEISHER: Did you do any head to head comparisons of this severity index model against some of the other more established ones?

MR. DeBRANTES: Well, we are doing it now, and Rand is actually conducting that study right now.

CO-CHAIR DUBOW: Okay. We are going to have a discussion of the measure itself as soon as we grab lunch. We are going to hear from Sean, because Sean is going to

give us his -- Is that okay, Sean? We are going to get lunch. Don't go away. You can get lunch, but don't leave. Don't leave.

So is 20 minutes adequate? Can we manage? We can eat in here. We are going to eat in here. Okay. Thank you.

(Whereupon, the foregoing matter went off the record at 1:05 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:27 p.m.

CO-CHAIR DUBOW: So the first thing I want to tell you is that we need a count. There is a dinner that has been arranged for here in the hotel at 6:15, and the van will pick you up, those of you who are staying at the hotel, at 7:30.

Could we have a count of those people who are going to be at the dinner, so we can make arrangements? Lee, this is a complete meal. Okay.

Do we know about tomorrow yet?

MS. BOSSLEY: No.

CO-CHAIR DUBOW: Okay. So plans for starting tomorrow are still not final. We are going to start at some point tomorrow, the earlier, the better, but we just don't know when.

Now we want to begin our discussion about this measure. So Sean is still having his conversation with Amita. We

are going to start with that, and we are also going to have -- We have a screen shot of the vote of the various Work Groups, Reva?

CO-CHAIR FLEISHER: It is on here.

CO-CHAIR FLEISHER: We have them. Okay. So let Sean give us his take on the methodology first. Okay?

DR. O'BRIEN: I was just trying to figure out what my take is, based on talking to Amita who was filling me in on very useful information.

I basically reviewed the measure based on what was provided in the measure submission, and then poking around in the web a lot and found an extensive amount of research looking at all kinds of aspects of the validity and reliability and the science underlying the method.

Most of the work that is there is dealing with modeling of the evidence in the form of case rates and estimation of cost. So I think my impression is that, although the

modeling is extensive, the modeling is focusing on cost rather than the frequency of PACs per se.

So that, to the extent that the models exist, they may have been optimized for one endpoint, and they are being applied to a different endpoint. So there would be a question of to what extent would results get there if you had a custom model specific to the endpoint of PACs, the binary yes/no occurrence, rather than the cost of an episode.

So I guess I will just continue, because I definitely have questions. Typically, when you are adjusting for an event rate, you want to adjust for factors that were present at the beginning -- you know, intrinsic risks to the patient that are present at the beginning of the care episode, and not adjust for factors that are influenced by the care providers that you are assessing.

My impression of the documentation

is that there are some variables in there that are actually things that are going on during the course of that episode. At least, that is a point that could be clarified. I may have been reading that incorrectly.

I guess it seemed like a dramatic amount of variation between the states, which is good news. I just would want to be careful to understand how the units that are being measured are entering their data, and could there be variations in the data collections, the number of diagnoses that are being filled in or present upon admission, indicators, anything like that that might be an alternative explanation for such very wide variation.

The mechanics of the severity adjustment: In my review I wrote that I wasn't clear that there was a severity adjustment for that PAC rate, that basically they are going to -- PAC rates are calculated as not relative values.

They say in their documentation, for example, a health plan would report that 60 percent of its plan members with a given chronic condition incurred PACs in the study time window. So they are just reporting absolute percentage like that. That has not been -- It doesn't sound like it is adjusting for the severity.

CO-CHAIR FLEISHER: Sean, we just heard they did. So is there --

DR. O'BRIEN: Well, so that is what I was trying to work through. To the extent that it is adjusted, I am not quite clear on the mechanics. So I can't really comment on that.

CO-CHAIR FLEISHER: But just before you finish that point, Francois, yes or no with regard to risk adjustment?

MR. DeBRANTES: Yes. So the severity adjustment model is done using dollars, absolutely, and it turns out a severity index, both at the individual patient

level and then rolls it up at the population level, which at that point allows you to determine the relative severity, for example, of a diabetic population versus any of the other chronic conditions.

You can look at the severity of sub-cohorts within the total population studied and determine the differences in the severity index and adjust the PAC rate for that. And we are, in fact, doing that, for example, with some payers in New Jersey and in other parts of the country.

The variation that exists in our developmental database has been replicated when we studied other databases, although again we see striking differences in rates of potentially avoidable complication from system to system and plan to plan, using the exact same methodology; and it is not, at least for us, explainable by -- We haven't been able to explain that from differences in documentation.

If it is there, it is likely to be standard noise across more than specifically focused in one location versus another, but unknown. That is a limitation of any claims dataset. You are going to have some amount of noise in it, and that is what we exclusively use for this measure, is claims data.

DR. O'BRIEN: I guess -- The mechanics of the severity adjustment aren't clear to me. I think what I maybe understand is that you -- Basically, each patient, you are going to have some weighted average of the risk factors that are present --

MR. DeBRANTES: That is right.

DR. O'BRIEN: For each factor that is present, you add on a certain amount and calculate that patient's severity score, and now you can, for a population, calculate the average severity score of that population or you can calculate the sum of the severity scores, and you can concur, okay, relative to some maybe benchmark population or rough edge

population, is this hospital, is this plan more severe or their case mix more severe compared to the benchmark population or less severe. You can quantify that by the ratio.

This cohort that we are interested in measuring has an average severity of 2 compared to the benchmark as one. So they are twice as severe as the benchmark.

Now how you take that result now and you see that this population that you are measuring has a PAC rate of 70 percent -- their severity is double the benchmark population. Now how do you take that 70 percent PAC rate and adjust it for the fact that their predicted -- their average severity score was twofold compared to the national population? To me, that it is not clear.

DR. RASTOGI: You just --

MR. DeBRANTES: You use the factor. Use the factor, and you apply the severity factor to the PAC percentage.

CO-CHAIR FLEISHER: Are we taking

-- Do we want to take questions right now or do you want to continue your critique?

MEMBER TURNER: So can we clarify just a touch more. So what goes into this is just dollars, though? The severity adjustment is just dollars?

MR. DeBRANTES: Is all kinds, yes.

MEMBER TURNER: Okay. And so if somebody for some reason has a physician who does lots of testing on them and does lots of MRIs, they are going to have more dollars.

MR. DeBRANTES: That has nothing to do with it.

MEMBER TURNER: Okay. So what dollars do you look at?

MR. DeBRANTES: Because it is not about the total dollars consumed by a patient. It is about their risk factors, and their risk factors are the types of comorbidities they might have based on, for example, the drug regimen or the types of office visits they have had.

So it is about risk factors of patients, not about --

MEMBER TURNER: But it is not disease diagnosis based. It is dollars attached to disease? I am still trying to get it straight.

MR. DeBRANTES: It is code based. I mean, I guess all that is relatively -- I guess, maybe not relatively well explained in the document, but it is -- For every patient population, you've got specific code sets that define the risk factors, and there is a comprehensive list of all the risk factors.

MEMBER TURNER: Okay. Diabetes and hypertension.

MR. DeBRANTES: That is correct.

MEMBER TURNER: Okay, I am getting it.

MR. DeBRANTES: So if you go through the spreadsheets, you've got a comprehensive list of all the risk factors for each population, and those risk factors are

calculated population by population.

So based on the relative profile of a given population, you are going to have a risk profile for that population and the individuals within that population, which is going to give you your severity index.

MEMBER TURNER: I missed the dollars part. That's the thing.

MEMBER ROSEN: So I understand what you are saying, are you saying that we then have this risk profile, and you look at the PACs, P-A-Cs that the patient has, and they will determine what the cost will be?

MR. DeBRANTES: Well, we certainly look at cost, but the measure as presented is counting the total number of potentially avoidable complications, not the dollars associated to them.

MEMBER ROSEN: Right.

CO-CHAIR FLEISHER: I guess where people are getting confused is we all -- many of us are used to an Alex Hauser or Charleson

or some sense of comorbidities. Are you using that or not? Is it just a simple question that can be --

MR. DeBRANTES: No.

DR. RASTOGI: The comorbidity index like Charleson index is approximately similar.

CO-CHAIR FLEISHER: Right.

DR. RASTOGI: Because these are -- Quite a few are outpatient, and we saw only six percent of the PACs was anything to do with hospital. Ninety-two percent-plus was on outpatient care. So we couldn't do Charleson's index.

MEMBER TURNER: I mean, there's tons of outpatient --

MR. DeBRANTES: Yes. So I am going to go back to -- our not-for-profit organization was primarily funded for the development of a program, and we decided specifically not to use any commercial application as a matter of policy, so that we

could put our work in the public domain at no cost.

CO-CHAIR DUBOW: Barbara.

MEMBER YAWN: Is there a clear distinction between a PAC and something you use for risk adjustment, because where I am concerned is could PACs be calculated in the risk adjustment and then for -- of course, you would expect them to add more PACs?

MR. DeBRANTES: That is an excellent question, and I think it is what distinguishes this approach from all the current episode approaches, is that we do not specifically exclude all potentially avoidable complications prior to looking at risk factors.

So risk factors are purely designed and evaluated on the typical services of the patients, excluding all potentially avoidable complications. Otherwise, you get into the circularity of --

MEMBER YAWN: You try to avoid

that.

MR. DeBRANTES: Exactly. Exactly. So risk factors are risk factors that exclude completely potentially avoidable complications.

MEMBER YAWN: Thank you.

CO-CHAIR DUBOW: Iver.

MEMBER JUSTER: So at the bottom of all this, would the intuitive idea be that you have two populations now, and one of them has a twice as high risk. So you would expect them to have twice as high PAC?

MR. DeBRANTES: That is correct.

MEMBER JUSTER: Okay.

CO-CHAIR DUBOW: Amy?

MEMBER ROSEN: A couple of comments. One is that it is difficult to separate out from a diagnosis code or something present on admissions is also a complication. There has been a lot in the literature on that, and that is why present on admission codes have been introduced in the

private sector.

I am just wondering if you have looked at that, because you really don't know if something is a complication or present on admission.

My big concern also is that how are the PACs determined? Was this by a clinical panel that determined whether or not a potentially avoidable complication was related to the index condition? How did you come up with this list of potentially avoidable complications, because there is a literature, you know, on this, starting with all these complication screening programs and some of the patient safety indicators from AHRQ are certainly important complications of care.

So there is one out there. AHRQ has also done preventable hospitalization.

MR. DeBRANTES: Absolutely.

MEMBER ROSEN: There are a lot of episode groupers.

My third point is that there are a lot of episode groupers out there. Have you checked your methodology comparing yourself to the reviews or the -- But it would be really important to see how you have conceptualized these components as your measure, whereas some others have been doing it all along.

MR. DeBRANTES: Sure.

MEMBER ROSEN: And what is the contribution of your measure?

MR. DeBRANTES: Sure. Well, none of the others use potentially avoidable complications. So I think that is a clear distinction, which is why we are here, and in fact, we are working with Ingenix now so that they can incorporate our definitions of potentially avoidable complications into the ETGs.

The issue of PLA applies mostly to the potentially avoidable complication measures that we are going to be looking at later this afternoon around MI, pneumonia,

and stroke. In our submission for those acute medical events, we do specify that conditions that are present on admission would not be counted as potentially avoidable complications for the reasons that you specified.

When reviewing the patients in chronic conditions, PLSA is a nonissue, because the ED visit itself is a potentially avoidable complication, and you would expect a patient who has diabetes admitted for an emergency department visit for, say, hypoglycemia to have a PLA diabetes. So it is not as applicable an issue on chronic care avoidable complications as it is on the inpatient ones and, certainly, for the inpatient ones we do exclude PLA for the reasons that you mentioned.

Then I think your other point about harmonization, I think, is what you were pointing at. Absolutely. We have incorporated in our definition of avoidable complications all of the existing ones to

date.

So again this afternoon when we go through MI, stroke and pneumonia, you will see that CMS defined hospital condition or defined PSIs -- all those are included as potentially avoidable complications. So we use those definitions and incorporate them.

We do take, and make no excuse for it -- quite the contrary -- We do take a very liberal view of potentially avoidable complications. So our list is far, far broader than what you will find at CMS, AHRQ, or anywhere else, and that is on purpose.

MEMBER JOHNSON: Francois, how did you establish that the relationship between risk and complications was linear? Then how do you assure that it is linear across all disease states when you are looking at multiple disease states? So you go up by a factor of two times risk. Does it equate to two times the complication 50 percent reduction and --

MR. DeBRANTES: Oh, you mean in the industry?

MEMBER JOHNSON: -- and is that consistent across everyone of your extrapolations when you look at multiple diseases?

MR. DeBRANTES: Yes. So don't know, and you know, when you do something like severity adjusting a rate of potentially avoidable complication, you have to make a decision.

In this instance, the decision that we made is, believe it or not, keep it simple. So is it linear? I don't think so. Again, the population that we studied is commercially insured.

So that is a more -- It certainly is a more homogeneous population than you would get if you took like, for example, an non-payer dataset, and you might come to some different conclusions in non-payer dataset.

In the commercial insured

population, I think it is -- you know, the linearity is more of a reasonable assumption, and is it perfect? No, by no means. I just don't -- You know, there is no such thing as perfection in severity adjustment. So you try to do something that is fair and reasonable, explainable, I think, to physicians and hospitals, so that when you hold them accountable, there is an understanding of the methodology.

CO-CHAIR FLEISHER: So can I ask: It sounds like it is not really risk adjusted. Sounds like you are going to actually present absolute rates with a severity index that the hospital could risk adjust, because it doesn't sound like you would be presenting risk adjusted rates, or am I missing something?

MR. DeBRANTES: No, you are not missing. You are correct.

CO-CHAIR FLEISHER: Okay.

MR. DeBRANTES: And back to the conversation this morning about the Minnesota

Community Measurement measure on diabetes which, by the way, Bridges to Excellence uses as its top level of performance in the country as we recognize physicians -- there is no severity adjustment.

The purpose is to count events. I want to relate a comment that we had during the discussion with the stroke PACs with the TAP. There was a comment about comas, in particular, and I am bringing it up, because of the 400-odd potentially avoidable complications, six of them were comas.

The point from the neurologists was that, you know, in many instances coma is unavoidable for patients with stroke. I asked a relatively basic question, which is -- not being a clinician myself, which is: Can you -- Are comas always on unavoidable for patients with stroke?

Of course, the answer is no. So we don't call these absolutely avertable complications for very specific reasons. We

call these potentially avoidable complications. The purpose is let's start counting these things. Let's start understanding the system that exists or doesn't exist around patients, and let's work to reduce them.

You know, we don't think that you can get to zero. None of us think we can get to zero, but we could probably get from 90 percent to 45 percent. I know that bothers a lot of people, but I think it bothers the patient a lot more when something happens.

DR. O'BRIEN: I feel like I have heard two different answers to the question of whether it is risk adjusted or not. So I am confused about that. But I would just say that, based on what is in the measure submission, which may be what we should go on, there is not written down the mechanics of actually taking an observed percentage and adjusting it up or down to account for the risk. It just indicates -- there is not

amounts and values to say the purpose of risk adjusting PAC rates.

There is a model that has been validated for the purpose of predicting costs and, to the extent that it is applied -- I know there is all kinds of statistical and nonstatistical reasons for decisions when you are developing a measure, but it seems like ideally developing a model that is specifically for this event would be the preferable way to do it.

This seems like kind of a non-standard approach to take, basically, a model that is for cost, and then multiply your observed PAC rate up or down by the ratio of this population's predicted cost relative to something else -- I think there's pitfalls with that approach based on where I have seen similar approaches in another context.

The question is, if you have extreme variation between the units in these predicted costs, which we may expect, to what

extent will you result maybe be driven by what you are getting on the denominator? Your predicted costs rather than observed results, I think, would be appropriately rated.

I assume that didn't make sense.

MR. DeBRANTES: No, no, your point is well taken. I think the decision is do you apply some severity adjustment to these rates of potentially avoidable complications or not? In this instance, we have taken the methodology we have developed around severity adjustment and said you can apply that methodology -- whether it is the best or not is clearly subject to opinion -- and you can apply that methodology to severity adjusting the PAC rate.

CO-CHAIR DUBOW: Okay. Any other discussion? Perhaps we can see how the sub-group voted on this measure.

MS. BOSSLEY: Can everyone read that or should we kind of summarize?

CO-CHAIR DUBOW: Why don't you do

that?

MS. BOSSLEY: Okay.

DR. BURSTIN: Tell them where it is, too.

MS. BOSSLEY: It starts first on page 83 of your PDF -- 82, actually. That is where the important section starts.

In general, for importance the four people who reviewed this measure thought that it did completely meet the importance criteria for the gap. There is evidence, and it does meet a priority.

For scientific acceptability -- and for anyone who was on the Work Group, feel free to jump in and add any comments. I am just going to provide high level.

For scientific acceptability, again you were asked to get in on the sub-criteria. So for the specifications and the reliability, the four felt that it completely met the criteria.

For validity, we had a split

between completely and partially. Then when you go down and look at exclusions, looking at the risk adjustment that was provided, meaning whether you can actually determine meaningful differences based on how it is specified, there of you felt it was completely, one partially, and also we had one minimally for meaningful differences.

Comparability, again four of you felt that it was completely met. Disparities, we had a mix between completely, partially, not at all, and did not apply.

Then for the next one, usability: This is where we are looking at harmonization, does the measure bring added value, is it understandable?

For understandable, two felt it completely met the sub-criteria. Two felt it partially met it.

Harmonization: Three of you felt that it completely did. One said Not Applicable, because there wasn't any other

measure other than what you have before you.

Then whether there was added value, all four felt it completely met the sub-criteria.

Then the last in feasibility, again we are looking at can the data be produced as a by-product of care? Is there electronic means to collect it?

Looking at exclusions, do you need additional data sources or not? Inaccuracies within the data and implementation? Pretty much everyone felt that it was completely meeting the criteria except for the inaccuracies piece, and that was two for completely and two for partially.

So very few ratings of minimal or not at all.

CO-CHAIR DUBOW: Is there anybody on the Sub-Work Group who is able to talk about the understandable piece of it under usability? Does anybody from the Sub-Group remember that particular?

I am just interested in why this is not considered understandable. I mean, what I always consider under this category is public reporting piece and how consumers will understand this, and this one feels to me so unbelievably understandable that it exceeds the grade that it could get, because --

MS. BOSSLEY: I am wondering if the first comment that says very useful with the managed care population for which there is an assigned primary care physician or group, less useful with insured population where there is not a PCP. I mean, Lee, I think you were on the group.

CO-CHAIR FLEISHER: Yes. Well, I am not sure about that first one, but I think when you put a rate starting at 89 percent and that it can be both preventable -- potentially preventable and potentially not preventable, I think that is one of the concerns people have.

if I remember some of the comments

from the other TAPs, it is, you know, if you get into an accident or certain other conditions -- certain conditions that you can't -- There are certain preventable -- certain PACs that may not be preventable, but you have grouped them into large buckets, and that is why is some concern, especially at an 89 percent rate. People would think that you could get to a zero percent rate, which I agree, that is why you have articulated that you might get from 85 to 45, but I think if it was released to the public, it may or may not be understandable in that regard, because of the wide capture of PACs.

MR. DeBRANTES: Can I respond to that? Again, I think this was discussed earlier in the Community Measurement effort where you can go on the website and see rates of 45 to one percent compliance, and in Minnesota people understand that. It might be shocking to many, but they do understand it.

I don't think -- It doesn't seem

to me as if everyone is thinking, oh, there numbers should all be at 100 percent, more than there is variation, and there is a big difference between 45 percent and one percent or 85 percent and 45 percent.

CO-CHAIR DUBOW: I would just add that these measures, when reported well, have some context, and there is some help to the user in understanding how to interpret the measure.

So you don't just slap up a bunch of numbers without providing some kind of guidance. In that context, I think that these kinds of measures are eminently understandable to the public who want to avoid complications. It really relates to the patient-centeredness part of this approach.

MEMBER JUSTER: As one committee members, I think my main concern was just -- which was not an easy thing for me to do -- was pretend I am an uninformed consumer, and I am having -- I am living in this world with

ever shortening sound bites that are -- you know, that the news programs are trying to get me to understand something in nine seconds.

So I am competing with this, and the things that I might -- and I am sure there will be some consumer testing, because I really like this measure, but I would need to know, for example, it really isn't expected to be 100 to zero, like everybody who has a heart problem should be taking a statin unless they can't. That should be -- There should be zero noncompliance there.

Here, I am being expected to understand that it won't reach zero. So then I am going to say, okay, well, if it shouldn't reach zero, what should it reach, because the whole thing about variation might just go right over my head.

Also, I would want to know that these were adjusted -- these took into account characteristics that people had before they were measured that they came into the

measurement period with, all sorts of things, three or four little sound bites that might be digestible.

Then for the more curious reader, I guess there would be more, but these considerations would be true for any public testing, I think.

MEMBER McNULTY: Sure. I thought it was completely useful. I was really comfortable with it, and to sort of make that point even more clear, I put it in front of my 86-year-old mother with chronic heart failure and asked her if she understood it, and she immediately understood what that measure meant, and it was very easy for a consumer.

I happen to recognize that first comment as being my comment. So I can speak to the other side of it, which is, when I look within the provider community, the ease with which it would be to move that number to drive toward perfection.

I happen to come from not a

system, a hospital system that employs relatively few physicians, and in this world that we are measuring here that really does take a group of people. It takes a system to manage these patients.

So when I look at the usability with my provider hat on, I realize that when I am going to get in my world is a constant push-pull between physicians and independent practice, emergency departments, specialists who are not aligned, and hospitals, all saying not my fault, I did my part.

That doesn't mean that it is not a good usable measure from a patient's point of view, a consumer point of view. It is very understandable. It is a very different measure and takes us to a level of accountability that we have not been pushed to before.

MEMBER YAWN: From a group of providers' perspective -- and I will not claim this as mine, but when I asked a little bit

about this concept, they said, as soon you say potentially avoidable complications, you have just employed all the attorneys in town.

That was their take on it, is until we have malpractice reform, doing this kind of thing is of concern to them. Now I am not saying that makes it right. I am just telling you that that was what -- and when we put it out for public comment, I expect to hear those comments of, if it was potentially avoidable, somebody should have avoided it and, by golly, I am going to sue, because it happened to me. So just another perspective.

CO-CHAIR DUBOW: Any other comments?

MEMBER DELLINGER: You could rephrase it as possibly avoidable. Seriously, I mean, you know, I do a lot of work in surgical infection, and we have a lot of very well proven process measures that we do infection risk, but never take it to zero.

So we say a surgical site

infection is potentially preventable if the right antibiotic wasn't given at the right time, if the patient wasn't kept warm, if blood sugar wasn't controlled perioperatively. But we have things we can really measure, and with administrative data you can't possibly do that.

This would feel better to me if it said possibly preventable complications.

CO-CHAIR FLEISHER: Have you seen some New England Journal papers' titles.

MEMBER YAWN: No, it is a legal word.

CO-CHAIR DUBOW: Okay. Well, we have -- Francois, I don't know if you want to entertain that.

MR. DeBRANTES: Well, just to mention that our Board Chair, Alice Gosfield, is a relatively well known health care lawyer who has represented physicians and hospital systems for a long, long time, and is robustly published.

I think it is an issue that we have actually debated extensively, and she feels very strongly that the use of potentially actually is a very good protection for physicians against aggressive attorneys.

CO-CHAIR DUBOW: Okay. I think we need to bring this to a close, because we need to move on. So, Reva, would you help us navigate the vote, please?

DR. WINKLER: Yes. As before, we need to vote on the criteria, the four main criteria as well as your final recommendation.

CO-CHAIR DUBOW: Excuse me. Is there anybody in the public who wants to make a comment? On the phone? Is there somebody?

DR. HALL: Bruce Hall from American College of Surgeons.

CO-CHAIR DUBOW: Okay.

DR. HALL: I am just looking again for the reliability and distinction between providers. I have asked those questions several times this morning, but I

just don't see that information presented.

MR. DeBRANTES: So is the question about --

CO-CHAIR DUBOW: Reliability among providers.

MR. DeBRANTES: Well, we can certainly tell where the avoidable complication came from, and this is a measure, as we say in the submission, that is designed not at the individual physician level. So I want to be clear about that and reiterate it. This is not an individual physician performance measurement.

MEMBER HOPKINS: Could you verify what is the unit of measurement?

MR. DeBRANTES: We think practice, certainly a medical group, hospital, a health system, health plan. But if you are talking about physicians, I think the lowest unit of accountability would be the practice.

CO-CHAIR DUBOW: This is a comment from George Isham who is the Medical Director

and Chief Health Officer of HealthPartners in Minneapolis. I am not going to read the whole thing, but he says:

"The comprehensive complications of care measure developed by Prometheus Payment as part of their payment reform model holds a promise to change the locus of quality accountability and stimulate the type of patient centeredness expressed in the IOM reports. Use of these measures does not necessarily need to be tied to payment, but they can be used as a performance measure on their own to ascertain effectiveness of transitions and coordination of care.

"Prometheus potentially avoidable complications encourage providers to look beyond what they do and engage them in the accountability of what happens to the patient. For example, cardiologists managing a patient with congestive heart failure are held accountable not only for the PACs related to the patient's CHF, but for PACs related to

comorbidities.

"HealthPartners believes that, by endorsing measures like Prometheus PACs, the National Quality Forum will help move the current provider focused health system into one that is patient centered, and that is why we support the endorsement of PACs as comprehensive complications of care as a performance measure."

Now, Reva, would you please walk us through the measure?

DR. WINKLER: Okay. As before, we will go through the four criteria, and then the recommendation.

So the first one is for this measure, the importance to measure and report. This is a yes/no vote.

So how many vote that this important to measure and report? Everybody, and how many everybodies are there? It is Sean who is missing.

MEMBER NEWCOMER: You missed one.

DR. WINKLER: Okay. So scientific acceptability of the measure properties. This is where we will vote completely, partially, minimally, or not at all, your assessment of how well this measure conforms to the criteria under scientific acceptability of this measure's properties.

All for completely? One, two, three, four, five.

Partially? Thirteen.

Minimally? Four.

Not at all? No. Do we add up right. Okay, that's 22. Great.

Now on the usability criteria, completely? Fifteen.

Partially? This is usability now, partially. Four.

Minimally? One.

Not at all?

You didn't vote. Okay, add one, all right. I am still missing one.

MEMBER NEWCOMER: I am a partial.

DR. WINKLER: Okay. So you are the partial. Okay. So I got 16 complete, five partial, one minimal. Okay.

The last one is feasibility. Completely? Fourteen.

Partially? Eight. That's it.

CO-CHAIR DUBOW: Now we have to vote up or down.

DR. WINKLER: Right.

CO-CHAIR DUBOW: So I guess it is just all those in favor of this measure?

DR. WINKLER: Correct, going forward.

CO-CHAIR DUBOW: Going forward for public comment and for endorsement.

DR. WINKLER: All in favor? Nineteen, I think.

No? Four?

CO-CHAIR DUBOW: Wait a minute. It doesn't add.

DR. WINKLER: No, it doesn't. Barbara changed her vote. Right? Okay. So

it is 18 yes, and four no.

CO-CHAIR DUBOW: Okay. So this measure is recommended by the Committee.

Okay, now we have two NCQA measures, and the measure -- Is Sue Miller -- I'm sorry -- Milner on the phone?

MS. MILNER: Hi. This is Sue. I am here.

CO-CHAIR DUBOW: Okay. Great. So let's start with the HOS measure.

MEMBER NEWCOMER: What number is that, please?

MS. MILNER: I'm sorry. Which one are you starting with? The health outcomes survey measure?

CO-CHAIR DUBOW: Yes, please. Lee, that is measure number --

MS. MILNER: Judy Ng? Judy, are you on the phone?

DR.NG: Yes, I am on the line.

CO-CHAIR DUBOW: What number is this? Number 6? Okay, sorry.

Do we want to -- Reva, do you want to tell us anything about this measure, or the measure developer?

DR. WINKLER: Well, I was going to say, the summary pretty much describes what is going on with this measure. This is a survey measure. It uses essentially the reference Rand Health Survey, the VR-12 as sort of an underlying instrument behind it.

This is a measure that has been used in the HEDIS program, I believe just in the Medicare patient population. There are scales that provide essentially two summary measures. One is the physical component summary score, and then the mental summary score. So there are sort of two results from the implementation of this survey measure.

CO-CHAIR DUBOW: I have a question about this measure. I have a recollection that this was once fielded in the fee-for-service population. I see that it is restricted to the Medicare Advantage

population.

Is there a reason? Is CMS here? I don't know. I know -- Hasn't this been used in fee-for-service Medicare before?

DR. NG: This is Judy from NCQA. It has been piloted in fee-for-service. I think it was about 10 years ago, and I think essentially what happened was it was found to be a bit too expensive to perform in fee-for-service, because --

CO-CHAIR DUBOW: We are having trouble hearing you. You are fading in and out.

DR. NG: Okay, hold on. Let me take off my speaker. Is this better? Okay.

Yes, it has been piloted in the fee-for-service study before. For a number of reasons, I believe mainly related to cost, CMS decided not to go ahead and leave it in the fee-for-service population and restrict it just to Medicare.

CO-CHAIR DUBOW: But there is

nothing, in and of itself, that would be peculiar to the Medicare Advantage population. Is that correct?

DR. NG: That is correct.

MEMBER HOPKINS: To further your point, it actually applies to chronic populations, and why would you restrict it anymore than that? This one doesn't have to tie to age. I don't understand why --

CO-CHAIR DUBOW: Well, it actually used to be called a health of seniors measure, but you know, I just -- If it is a cost issue for the implementer, that doesn't speak to the measure properties, and I just wondered about that. Is there discussion?

Do the folks from NCQA want to say anything about this measure? Would you like to add anything to what Reva mentioned? David?

MEMBER JOHNSON: I just had a question. As with any survey, it is subject to who fills it out, and what is the

anticipation of how this is going to be used? You give it to a patient, and the people that are happy are going to fill it out, and the people that are very angry are going to fill it out, but the vast majority of people are going to say I don't need to do this.

CO-CHAIR DUBOW: There is some experience with this measure already. Can you tell us a little bit about that? This measure is used in Medicare, in the Medicare Advantage program already.

MEMBER JOHNSON: Again, I am just subject to a lot of surveys, and it is the people who fill them out have either one extreme or the other. It is the in-betweens that are the majority that typically say I don't need to do this.

CO-CHAIR DUBOW: Do we have data on the response?

DR. NG: Yes. The response has gone from 8 over 60 to 80 percent.

CO-CHAIR DUBOW: You know, CMS --

Well, I guess CMS. NCQA drives other measures from this survey. So this, in addition to being a functional status measure, also has embedded in the survey instrument -- I don't know -- flu and a couple of other measures that come out of this. It is in here somewhere.

DR. NG: The measure covers osteoporosis -- and there are a number of items on comorbidities.

CO-CHAIR DUBOW: And that is probably an aside, because what we are really considering here is the functional status measure.

DR. NG: There's other measures, I believe, are endorsed separately.

CO-CHAIR DUBOW: Yes. Right.

MEMBER JEWELL: Right. So this may just be my mental fatigue from having read so much over the last several days of this material, but how -- What are the score cut points for deciding better, worse?

I am not familiar with the FS-36 in its original form. So I know what the population norms are. I know different population important differences that distinguish between better, no different, etcetera. So is that in here, and I just missed it or can you provide some clarification about how do you decide who is better and who is not?

DR. NG: The better or worse things actually are based on national norms. I think at this moment they are the norms for the 1998 U.S. general population for that particular age group.

MEMBER JEWELL: Were the instruments in this version or with the original SF-36? Maybe that is why I am confused?

DR. NG: I think it was the basic norm as well as -- They have been working here with a 12-item instrument using this version, and I believe it would be 1991.

MEMBER DELLINGER: My only comment on this is, if you look on page 9 of the document that we were given here, it says that out of 187 MAOs, two had mental health better than expected -- that is one percent; 10 worse than expected -- that is five percent; zero had physical health either better or worse than expected.

So this distinguishes nothing. This is a useless instrument.

DR. NG: I think part of what is happening with that is the way the risk adjustment is being done. They might actually be risk adjusting for a lot of the factors, and that is what we are actually looking into right now.

DR. TURNER: I think it is excellent information to have, with the onus of understanding the case mix of your population, but I guess I have been always wondering about the research that shows a lot of these measures are meetable by the care

that you give.

I mean just that it is a small portion of how your mental health is. What is going on with your health care? There are so many other factors in terms of where you live and your socioeconomic status.

So I am just curious what you would use this for or is it just to inform your study and your sites about the case mix that they take care of?

CO-CHAIR DUBOW: Sue, do you want to respond to that, or maybe CMS would respond to it. Come to the table, please.

DR. HALIM: This is Shaheen Halim from CMS. The mental health score and the physical health scores are actually used in some of our health plan performance metrics that are shown on the Medicare options compare site. So I just wanted to point out that use. That is -- It is being publicly reported in that website.

CO-CHAIR DUBOW: I thought the

question was what could be done by the plan to improve the scores.

DR. HALIM: Oh, I see.

MEMBER TURNER: Do you have any data that getting those data allow the plan to actually change the scores of their patient populations, because that is -- because they are usually so multi-factorial -- what makes your mental function, etcetera -- that it is a little piece of the pie.

DR. HALIM: Right. I don't have that information. Perhaps NCQA can comment on how it is being used in quality improvement activities.

CO-CHAIR FLEISHER: So can I just ask. Does CMS use it as a -- just reporting it or as a performance measure, because if NQF chooses to endorse it, then it can go for a performance measure. If NQF doesn't choose to endorse something, given the current -- my understanding of what is in the bill and the language, then what happens?

DR. HALIM: I can't speak to that, but I do know that it is used as part of a composite -- a set of composite measures that are shown on Medicare --

CO-CHAIR FLEISHER: Are those composite measures endorsed?

CO-CHAIR DUBOW: That is what CMS does. But I think that nothing compels CMS to use NQF endorsed measures, although the legislation does -- Well --

DR. BURSTIN: The current status of it is that NQF is they will have to look toward NQF for standards when they are available. I think the issue here has not endorsed any functional status measures to date. So they have not been available. So they will have to use others.

This has been viewed as a -- Actually, as part of the work that the Measure Prioritization Committee of doing, functional status rose to the top of the list as being a pretty significant measurement gap.

CO-CHAIR DUBOW: Right. The new legislation does say that the Secretary should give preference to NQF endorsed measures.

MEMBER HOPKINS: And says functional status is important.

CO-CHAIR DUBOW: Yes, is a key issue. I would point out that in the long history of this measure, and it has been around for a really, really long time, that the HOS group at CMS has tried to respond to that question that you have raised by publishing guidance on what plans could do to improve the functional status of their members.

So there is some literature on this by way of guidance, but obviously, we see not a whole lot of variation in the reporting performance, perhaps because of the way the measure is risk adjusted.

MEMBER HOPKINS: Is somebody addressing that issue?

CO-CHAIR DUBOW: I heard Sue say

that they are looking at the risk adjustment. Is that right, Sue?

DR. NG: This is Judy speaking.

CO-CHAIR DUBOW: I'm sorry, Judy.

DR. NG: That is correct. We are looking at it right now, for the same reasons that the outlier is showing up.

CO-CHAIR FLEISHER: But we have to vote on the current risk adjustment.

CO-CHAIR DUBOW: I was just going to ask, what is the timing on your reevaluation?

DR. NG: Until, I believe it is, the end of the summer, but it is an issue that we could revisit even at that point.

MEMBER HOPKINS: Can this committee carry something over that long? I am curious.

DR. WINKLER: It is not going to match the timeline of this project. The question is, is there another avenue within NQF to potentially look at the measure and,

you know -- unclear at this point, but probably will be.

DR. BURSTIN: We would also have to go back to NCQA and really find out. Even if the risk adjustment, the work, is done by July, when will we actually have results. These they will bring to us. We always have the option of doing an ad hoc review if something changes significantly, if it is endorsed.

CO-CHAIR FLEISHER: What if it is not endorsed?

DR. BURSTIN Then they would have to wait for another opportunity to resubmit it.

DR. NG: It is possible that it could be done earlier than that, considerably less time.

CO-CHAIR DUBOW: Dianne.

DR. JEWELL: I am still struggling with trying to figure out what this measure is for. I hear where everybody is interested in

having functional status measures, but I don't know what this measure is supposed to tell me.

DR. HERMAN: I would second that. If it is an outcomes measure, there should be something that we can do to change it, and I haven't heard anything that shows that you can do anything to change it. So if it is supposed to be an outcomes measure, there has to be kind of the before and the after.

DR. JEWELL: Right. And again, maybe I just am not seeing all the detail that is in front of me, but as it is described right now, it sounds to me more just like a status check, like how is the health of the population that we happen to be looking in on.

CO-CHAIR DUBOW: No. This is a two-year measure. This follows a cohort over two years, and it is to change from expected. It is a two-year --

DR. JEWELL: Okay. So then if that is what I understood the first time, I am wondering how valid it is to talk about the

change in expected in a score like this, when you are talking about norms that are validated with a different instrument.

DR. HERMAN: Or if there is no change over the two-year period.

DR. JEWELL: What does it mean? Yes.

CO-CHAIR DUBOW: Judy? Sean, do you have any insight?

DR. O'BRIEN: I think the outcome of the measure is reporting patients that don't deteriorate, that maintain their status or improve. So, basically, they are using the same measure baseline and two years later comparing them.

MEMBER NEWCOMER: And it is that a function of health care or a function of --

CO-CHAIR DUBOW: Barbara?

MEMBER TURNER: Well, but there are things that you can do to improve the mental health of the community, like perhaps recognizing depression and beginning to treat

it, which we do very badly as a health care sort of a system -- I guess we are.

So there are some things like that that can be done, but is this the measure to do it when they have such a small change and so few outliers. It bothers me that this is not a good measure for assessing how we are currently doing in giving us room to improve.

So that is my biggest concern.

CO-CHAIR DUBOW: So I heard some interest in deferring -- is that the right word? -- to hear how NCQA proposes to modify the risk adjustment, to see whether we could see more discrimination?

DR. TURNER: I think that is a great idea, and I also suggest, if they have any data that helps us understand that health care delivery has something to do with that or whether we can parse out our contributions that we should be responsible for and if we make a difference; because I think it is a great thing.

CO-CHAIR DUBOW: Wait a minute. We need some guidance from Reva and Helen about whether we can -- what we can do to move that sentiment into reality.

DR. BURSTIN: It sounds like the first thing we should do is just ask for additional clarification from NCQA, the additional analysis they can provide. When is the schedule for testing? And I think we will have to -- I think we are asking people on the phone to make sort of off the cuff assessments of when things will be ready. I really want to go back to the NCQA leadership and be able to do that.

I think it would be helpful to get a full set of what the issues are. The issue Barbara just raised is a complicated one. Oftentimes we don't always know exactly what health care interventions affect outcomes, and yet if we think they are worthy and important outcomes, we still go ahead and put them through.

So I think that may be a harder lift for NCQA to really respond to.

CO-CHAIR FLEISHER: So, Helen, if we vote no, will we get a chance to relook at this. If we vote yes conditionally, and they don't -- We can defer?

CO-CHAIR DUBOW: We are not asking -- What we are exploring here is the possibility of not taking a vote, so that we can go back and ask NCQA a series of questions to give us some better clarity. Based on those answers, we could vote by email. Okay?

So let's just take a couple of minutes to flesh out what we want to know. One of the things is that we want to know about their plans to reexamine the current risk adjustment method to see whether they are going to come up with something that will be more discriminating with respect to the results. That is the first item I hear. Is that right? Okay.

The second is some data around--

MEMBER TURNER: Any influence that say depression interventions or something on a large population, anything that we can do.

CO-CHAIR DUBOW: Yes. This is not a new measure. This has been used. I mean, we have those data, by the way. You know they went back -- I can't remember.

It goes all the way back to the early 2000s or the late '90s, I think. So we saw some stuff here about change in performance, and some of the literature that they provided us also discussed that, if you click on some of those links. They have a whole website. So that is one thing. Dianne, you want to add to the list?

MEMBER JEWELL: Yes. I just need some direction about what constitutes meaningful change, so that you can fall into one of these categories or not, and how that has been determined and validated. That is really my point about validation. I didn't articulate very well the last time. Even

standard error change.

DR. O'BRIEN: They have extensive documentation on the web on this, and it was all provided with the submission.

MEMBER JEWELL: Okay. So if I need to, I will do that.

CO-CHAIR DUBOW: All right. So we have Dianne's point.

DR. O'BRIEN: Can I add a couple to the list? Things I noticed is that one of the questions that I was asked to address to my view is whether the risk model adjusts for factors that reflect disparities in care.

This is a measure that does adjust for socioeconomic status and race and goes back. So it definitely does. So whether that is right or wrong or if it is not exactly consistent with the current NQF criteria for evaluating risk adjustment measures, it is a relatively limited set of risk factor adjustments.

So for adjusting, there is a

mortality model. There is an MCS model and PCS model. That's the mental and the physical component scores. The two component score models basically only adjust for socioeconomic variables. They do not adjust for baseline measurements -- you know, they don't adjust for your baseline MCS score or any other factors and may be associated with the likelihood you will be able to maintain your current health status.

You know, the methodology was extensively and rigorously tested from all kinds of perspectives. I mean, it is clear that this is a long history, lots of publications, and really a lot of work went into it.

Looking at it, I couldn't tell whether fit of the models they are proposing were assessed in terms of calibration and looking at calibration within the subgroups.

The last point was that, in terms of this discriminating performance, I saw the

same results that you raised. I thought the PCS was less discriminating. I mean, there were outliers for the -- actually, for the physical. One appeared to more underpowered than the other.

CO-CHAIR DUBOW: More discrimination, but not a lot. That has been pretty consistent over the years as well.

MEMBER JUSTER: Since we are measuring compared to themselves, is the role of risk adjustment when the outcome is a difference between time one and time two may not --

CO-CHAIR DUBOW: That is the point.

MEMBER JUSTER: And left to myself, I have a chronic disease, I am simply going to deteriorate.

CO-CHAIR DUBOW: Do you want to add something to that, Judy?

DR. NG: It is --

CO-CHAIR DUBOW: We cannot hear

you. Okay. I think it adjusts itself, because it is the same person.

MEMBER AMARASINGHAM: I am not sure of that, though. Let me just make sure I understand that. I mean, you can still risk adjust for the likelihood of something happening to a patient. So for example, you can risk adjust for the likelihood of a readmission in the future for a patient. It's still the same patient.

MEMBER JUSTER: I am not saying don't risk adjust. I am just saying --

MEMBER AMARASINGHAM: But I think risk adjustment is still important. Don't take risk adjustment out of the table.

CO-CHAIR DUBOW: Amy?

MEMBER ROSEN: I just want to raise maybe what Sean was thinking about, too, is kind of the clumping of using the summary components, summary scores for MCS and PCS and whether that is important in thinking about the lack of variation in the outcomes, whether

looking more specifically at some of the domains of the summary scores might be more effective in picking up more variation. Just as a thought statistically in terms of moving us more forward.

CO-CHAIR DUBOW: I would like to just add -- I mean, this is almost a rhetorical question about the issue of restricting this to the Medicare Advantage population and whether there is something intrinsic about this particular measure that justifies that, excluding the fee-for-service population.

I expect the answer to be no, but I would just like it on the record. So are there any other questions? It sounds as though there is consensus around deferring a decision on this measure, on the HOS measure, until we get some feedback from NCQA, the measure developer. Okay.

DR. PAGE: Joyce, this is Karen Page, NQF.

CO-CHAIR DUBOW: Hi, Karen.

DR. PAGE: I just want to address a question that came up about whether you need to risk adjust if you are using -- or comparing difference to the patient's own baseline.

The question that comes up is that, depending on what your baseline is, there may be different opportunities or probability of improvement. So if that is the case, so say someone -- whether you are the higher end, to begin with, and you have greater chance of improving or if you are the lower end and you have greater chance of improving, the idea is that there is a different mix of patients that is starting at the different levels. There is a variable probability of changing, but in risk adjustment it is something to at least consider.

CO-CHAIR DUBOW: Thank you, Karen.

MEMBER HOPKINS: My process

question was is that going to happen in the time frame of this committee?

CO-CHAIR DUBOW: Is what going to happen?

MEMBER HOPKINS: Whatever is taking place. We are deferring decision, but I would like --

DR. WINKLER: We can ask NCQA and get a response within the time frame. What that response will then set you up to do will be the next step. You may not be able to act or do anything within the time frame remaining of the project, depending on what the response is.

MEMBER HOPKINS: So I am trying to figure out how this gibes with the decision we made on the Minnesota measure where it was conditionally approved. Seems like we are not treating things the same way.

DR. BURSTIN: It sounds like people don't feel like we have enough even to make that decision or even say what the

conditions are. There is enough outstanding questions that I have the sense people aren't ready to make that choice. If people feel ready to make that choice, that's another option, but I think the bigger issue is when is this testing going to be done? Would you want to see the updated tested measure before you make that decision?

CO-CHAIR DUBOW: Dianne.

MEMBER JEWELL: Can I just clarify? This measure -- I infer references I have seen in the reference list that this measure has been around for a while. Has it been around for a while in the Medicare Advantage?

CO-CHAIR DUBOW: Yes.

MEMBER JEWELL: But the reliability testing that is reported is only with the veteran's group? Did I understand that properly?

CO-CHAIR DUBOW: Is that correct, NCQA?

MEMBER JEWELL: What is in the reliability section, at least in the measure application form, refers to extensive testing in the Veterans Affairs study with that population, and I just am not clear the extent of the testing in other groups.

DR. PAGE: I believe this has been tested in both veterans and elderly groups.

CO-CHAIR DUBOW: Thank you.

MEMBER HOPKINS: Joyce, what I am struggling with is the core of this survey is SF-36,probably the most widely tested survey instrument in the world. So if we get through this process and don't even have NQF endorsement of SF-36, something is wrong.

MEMBER JEWELL: This isn't the SF-36.

MEMBER HOPKINS: It is embedded in this thing.

MEMBER JEWELL: But it is not the SF-36.

CO-CHAIR DUBOW: Okay. You know

what. Let's take a vote to defer, because we need to move on, everybody. We are late. We are way behind. So I think, as Lee suggests, that we take a vote.

All those in favor of deferring consideration of this measure until we have answers to the questions we just identified, please raise your hand.

Okay. So we all agree, and we will defer consideration. We will be hearing from NQF staff. Keep an eye out for the email.

We now have the last cross-cutting measure for our consideration, and that is care for older adults, advance care planning, functional status assessment, pain screen. Again, it is an NCQA measure.

Reva, do you want to introduce us to the measure?

DR. WINKLER: Sure. This is measure 007. It is care for older adults. So percentage of adults 65 years and older who

receive the following during a measurement year: advance care planning, functional status assessment, and pain screening. Each of these are reported individually, though they are part of this measure.

MEMBER HOPKINS: Page reference is 25.

CO-CHAIR FLEISHER: Thank you.

CO-CHAIR DUBOW: Yes, and it is 0T2-007-09.

DR. WINKLER: This is not a composite measure. It has multiple parts embedded in it -- or it is not submitted as a composite measure. Let me put it that way, and we have endorsed similar measures that are sort of multi-part, if you will, like this.

CO-CHAIR FLEISHER: So there would be one person who voted? Is that --

DR. WINKLER: Yes. Not a lot of participation for the group that looked at this one. So whoever was the one person, thank you very much for stepping up.

MEMBER HOPKINS: Reva, this one raises a fundamental question. How did it get through the screen for outcomes measures? It is not an outcome measure in any sense of the word.

DR. WINKLER: What screening are you referring to exactly?

MEMBER HOPKINS: I thought this was the Patient Outcomes Steering Committee. Wasn't that the first slide that you showed? Where does it fit on that slide? This is a total process measure.

DR. WINKLER: And the Steering Committee is welcome to make that determination and act that way.

MEMBER NEWCOMER: So, Mr. Chairman, let's emulate the lawyers and ask for a summary judgment to dismiss, because it really isn't an outcomes measure.

CO-CHAIR DUBOW: Is there anybody who wants to entertain this measure?

MEMBER TURNER: I had one

question. I mean, I think we still have to consider process measures, because that moves the bar in the right direction, but -- am I wrong? Advance care planning? I am reading the wrong thing. Am I?

CO-CHAIR DUBOW: No, you are not. You are not.

MEMBER TURNER: I think they are processes. I think process measures can be outcome measures.

CO-CHAIR DUBOW: Excuse me. This is an important discussion.

MEMBER HOPKINS: Instead of saying I assess the functional status, report what it is.

MEMBER TURNER: Meaning what was the pain, not pain screening. What was the pain?

MEMBER HAUGEN: From a patient standpoint, the fact you do something isn't meaningful. It is what you do with that information, and then do I improve it. So

just the fact I do things -- I just don't think that is even comes close to accountability.

MEMBER PINDOLIA: You know, there is NQF measure number 0553 for medication review. That is why they didn't include it in the three.

I did vote, but I think I wrote it after the due date. So I did vote, but I guess it didn't get counted, because I was a little late. Sorry about that. But it says that the 0553 has already been NQF endorsed, and that is why they didn't include that in here, and they included the other three senior outcome measures.

Is there any data? Is 0553 any different than these or what they are looking for? It is on page -- It is under 3(c), distinctive or additive values in other NQF endorsed measures.

CO-CHAIR FLEISHER: So, Helen, within the context of this discussion, so if

the committee -- Somebody has proposed a potential measure. So the question is, if we say no, do we say no, because we don't believe it should be endorsed or, no -- do we also have a right to say, no, it is not an outcome measure and, therefore, it should not be brought forward? You are winding your eye. So that means --

DR. BURSTIN: It is very late in the process, and I think our understanding was all these measures were looked upon against this list of what we decided up front were broad topical areas that could -- Some of these aren't the classic outcomes. You had a fairly broad list up front of what you considered outcome measures.

So we made that initial assessment, I thought, with you guys, actually, to specifically include these kinds of -- that this measure was included, because it actually fit patient experience.

MEMBER HOPKINS: But back to that

discussion, we weren't talking about somebody assessed those things. We were talking about the measure was measuring those things. That is the distinction I would make.

MEMBER JEWELL: The comparable conversation that we had in the fall related to another measure, was the gate velocity measure that I had raised a question about. The response I got was that in its current specification, because it was the question, did the physical therapist assess gate speed, yes or no, that that was by definition a process measure, and that that wasn't relevant to this conversation, and that in order to make it relevant, it would need to be respecified to reflect some set of values against which you would hope the patient would match, like the A1c measures.

So using that logic, I concur that this --

MEMBER TURNER: So with that analogy, what would be the advance care

planning outcome? They stayed alive or they didn't stay alive or they -- I am just saying that you do want to have -- You do want to have the process evaluated.

MEMBER JEWELL: And I don't disagree with you philosophically. I think it is a matter of where does it best belong. I guess I do have a little bit of a concern about integrity of process by virtue of other measure submitters.

I think we need to be as clear as we can be that the rules are applied the same way all the time. So if other groups thought they had measures -- and I speak to this more from the bone and joint TAP. We got no measures. I can imagine there was a whole wealth of process measures that would have been relevant in the same logic that you are arguing.

So for me, it is both a consistency of approach issue as well as just a definition issue.

CO-CHAIR DUBOW: David raised that point earlier this morning in a sidebar conversation, and I think that argument has merit. But I do think we need to be, to Barbara's point, very clear that it doesn't fit into our definition, because that is the standard that we will be judged by, and any action we take needs to be justified on the basis of the fact that it doesn't meet our definition.

This is the definition that we advertised, and I suppose if we decide that it doesn't, then the measure is not considered but without prejudice.

MEMBER BECKER: So, Joyce, just a question. So I think these are important things, whether they are process or outcome. So if we decide not to go forward, is there a place where these get posited so they can get accepted -- reviewed, accepted, not accepted, because I think advance care planning is an important thing to do. You know, if we just

pocket veto these things and they fall into an abyss, then I don't think we are doing --

CO-CHAIR DUBOW: Is there a shared decision making -- Nothing? Is there any other place to give this --

CO-CHAIR FLEISHER: So actually, getting back to Barbara's comment, if advance care planning is considered an outcome, if they have a plan, then it doesn't matter if the others are process measures, from the previous discussion, as long as one of the components is an outcome.

So the question is -- and I think it is a great question that Barbara asked -- Is there anything that -- If you create a plan, is that an outcome versus a process.

CO-CHAIR DUBOW: Having a plan.

CO-CHAIR FLEISHER: Having a plan.

MEMBER AMARASINGHAM: But I think the reason to have a plan for the outcome is that your end of life is better. Now we don't know -- how that is defined is very murky, but

that is the whole reason for advance care planning. I still think it is a process measure.

CO-CHAIR DUBOW: No, it is having a plan that reflects your preferences.

MEMBER AMARASINGHAM: Right. So that decisions can be better made at the end of life.

CO-CHAIR DUBOW: No. That reflects your preferences.

MEMBER AMARASINGHAM: For the end of life.

MEMBER JEWELL: Well, by saying it reflects your preferences, to me that sounds like an intermediate outcome. That is not how this measure is specified. So then we are respecifying the measure on behalf of the developer. So that is a whole 'nother issue.

MEMBER ROSEN: Only because it is -- the advance care planning is specified with CPT codes. So we are not really asking the patient what happened. We are looking at the

data.

MEMBER JOHNSON: Can we get back to the slide? Just put the slide back up, the one we just had, what we are charged to do.

There are a couple of circumstances here that we need to consider. One, this was accepted for this task force to review. So it is -- It got into the queue where other people would have not had these process measures maybe go forward. Is that right?

CO-CHAIR FLEISHER: Reva, did anything get rejected by the staff to say it was not -- it would not be reviewed? So if somebody submitted -- because if not, then that is a dangerous statement to make. So we should just have clarification. Did staff perform triage?

DR. WINKLER: Well, the problem is, staff had the same discussion you are having, with a variety of opinions, actually, and applying that as criteria is harder than

you think. So the default was to keep it rather than let it -- and let the Steering Committee make that decision.

MEMBER JOHNSON: So nothing was turned down? My point is just that --

DR. WINKLER: It was just one or -- you know --

MEMBER JOHNSON: The second point is: I am 100 percent, this is a process measure, but if you look up and read what we said in bullet 1, patient function, symptoms, health related quality, I think we are really trying to talk about changes in, rather than measurement of.

If you just measure something, and in the first bullet I think that is still a process measure. It is not an outcome. You haven't defined a change, which is what an outcome is.

So I think, if I were submitting this measure and I read your first bullet point, I would say, well, we fit right into

that.

CO-CHAIR DUBOW: I want to ask the measure developer. NCQA, can you tell us why you submitted this measure as an outcomes measure, please? It goes to the question that you raised, David.

DR. PAGE: Yes. I believe that we had discussions with staff and felt that, you know, in terms of measure call and where this measure might be most appropriate, you know, this was where we ended up.

CO-CHAIR DUBOW: Okay, thank you.

If there is no further discussion, I think we need to call the vote, and it seems to me that we ought to have the vote on the basis of whether we consider this -- before we vote on -- Well, we could do it up or down, but I think we should vote on whether we consider this a process measure, because it is out of scope. Right.

MEMBER PINDOLIA: Hold on. Before we vote, I think my question still hasn't been

answered. There's four components for senior care of what they are trying to do. They only included three, because one of them was already NQF endorsed, and that was the medication review, number 0553 that they put in there.

Is that considered an outcome measure, because if that was, then you can's say these three aren't. But if that wasn't, then maybe that is what those should go to, whatever those are called.

DR. PAGE: I believe that was endorsed under a separate measure development call.

CO-CHAIR DUBOW: It was.

DR. PAGE: And perhaps Helen knows exactly what.

CO-CHAIR DUBOW: But that doesn't matter, because the scope of those projects was different. The criterion in the other project wasn't was it outcome or process.

MEMBER PINDOLIA: That is what I

wanted to know. Was it or not?

CO-CHAIR DUBOW: No, it wasn't.

So now we are going to vote. All those who believe this measure is in scope for this committee, please raise your hands.

Okay. Then out of scope?

CO-CHAIR FLEISHER: In or out of scope. Out of scope is now. So it is 21.

CO-CHAIR DUBOW: So I think this does not indicate anybody's preference or opinion about the measure itself.

DR. BURSTIN: Let's try to find a home for it, so it doesn't fall into an abyss.

CO-CHAIR DUBOW: Right.

MEMBER AMARASINGHAM: I guess one question is: Can our committee make a recommendation to NQF, because I think, even when we were thinking about this last fall and we were going out and talking to other methodologists, one of the things I mentioned to them is we are not looking for process measures. But that is equally vitally

important, and there should be some forum or, hopefully, there will be a forum at NQF where that is considered.

Now, of course, advance care planning is critical.

CO-CHAIR DUBOW: Right. This actually fits into one of the six NPP priorities, as a matter of fact.

DR. BURSTIN: We are planning to have a palliative care project beginning in November. So at the latest, it is six months away.

DR. O'BRIEN: I was wondering about the Brandeis CMS evaluation of management for heart failure, MI, and pneumonia. Those are arguably processes of care.

CO-CHAIR DUBOW: Okay. So we have now completed our consideration of the cross-cutting measures, and we are now going to infectious disease.

CO-CHAIR FLEISHER: Okay. So we

have ordered joe for the entire panel, could do jumping jacks to get us back on track. Do people want to take two minutes to stretch? Why don't we do that. So take two minutes to stretch while we get the next set of measures up, and we will start.

(Whereupon, the foregoing matter went off the record at 2:52 p.m. and went back on the record at 2:57 p.m.)

CO-CHAIR FLEISHER: Okay. We actually will get Francois back to the table. This is not you? Okay. No, no,no. We are going to start with your measure.

CO-CHAIR DUBOW: Under ID, pneumonia.

CO-CHAIR FLEISHER: So, you know, I am used to surgeons asking docs, you have a 7:45 heart start time, and I am used to calling them at 7:15 and asking where they are.

So we are going to start. So, B.J., we would like to start. So we have

already actually had a lot of these discussions. I want to keep this going.

DR. WINKLER: Just by way of introduction to this set of measures, hopefully, we can be a little bit efficient, because one measure is very similar to the measure we talked about earlier today about the PACs, and Francois is back to talk about that.

The next three measures for ID are very similar to a heart failure, AMI measures from Brandeis that we already talked about in March. So the methods are the same. The issues should be very much the same. So, hopefully, we can perhaps be a little more efficient in our conversation, without redoing things we have already gone over and over again.

So the first measure we are going to talk about -- and we need to go down to 22, Helen -- is the proportion of pneumonia patients that have potentially avoidable

complications during the index day or the 30-day post-discharge measure.

So this is brought to you from Francois and company, who created a whole suite of measures. This is a measure focusing in on patients who are hospitalized with pneumonia who then have -- again, same methodology, identifying the PACs -- either during their hospitalization or within the 30-day time frame immediately after hospitalization. It is measure 22.

CO-CHAIR FLEISHER: But the difference is there was one --

CO-CHAIR DUBOW: It is measure 13.

DR. WINKLER: Oh, you are right.

CO-CHAIR FLEISHER: So we have moved down from the one year to a 30-day, an important point.

DR. WINKLER: Right, and it is focusing in on patients whose primary discharge diagnosis was pneumonia. Correct, Francois?

MR. DeBRANTES: That is right. So primary discharge diagnosis is pneumonia and potentially avoidable complications, just like in the definitions for chronic illness, include readmissions or ED visits, encounters related to the pneumonia. That would be Type 1; Type 2 related to comorbid conditions; type 3 related to patient safety issues.

When you look at the distribution of the potentially avoidable complications for this type of measure, it is more concentrated around complications that occur related to the index condition and patient safety issues, mostly stuff that happens during the hospital stay.

There was a fair amount of discussion during the TAP for pneumonia, and I have to admit, not being a physician and Amita not being around, I was not able to answer a lot of their clinical questions, and I also clearly was not very articulate in my answers; because it didn't seem as if anyone

on the TAP actually understood what I was talking about.

CO-CHAIR FLEISHER: Barbara, were you the Chair?

MEMBER YAWN: No. I was on pulmonary, but this TAP is infectious disease.

CO-CHAIR FLEISHER: Yes, this was under infectious disease.

MEMBER YAWN: I don't know why, but it was.

CO-CHAIR FLEISHER: Right.

MR. DeBRANTES: So, for example, you know, we do specify -- I tried to explain a couple of times -- that if a patient comes in and has something present on admission -- we had that conversation earlier -- then, obviously, it is not going to be counted as a potentially avoidable complication.

We got into circular discussions around the severity adjustment, not dissimilar to what we had earlier.

CO-CHAIR FLEISHER: Francois, why

don't we get the TAP comments, ask for more comment form the TAP's perspective, the concerns about this measure or what they thought about this measure.

MEMBER DELLINGER: I think we had trouble understanding a lot of it, and the issues -- Francois, it helps me reflecting reasonably well the conversation -- and some of them are put up in the material here on page 71.

There was concern over -- For instance, thoracentesis was considered a PAC, and yet it is indicated if a patient has pleural fluid and the question of empyema. There are a lot of things included as PACs that seemed to us like necessary components of care.

CO-CHAIR FLEISHER: Amita, do you want to address?

DR. RASTOGI: So being a thoracic surgeon, I agree with you that thoracentesis is an important part of care, but empyema

itself should not happen, and sometimes the codes are picked up by empyema, and sometimes it is picked up by the procedure code.

The idea is that we can't avoid all empyema after pneumonia, but these are all potentially avoidable complications that we are talking about, and we want to restrict the number of empyemas that happen, just like we want to restrict the number of dates that happens in patients who are hospitalized.

So all these systems are potentially avoidable, and you are not expecting a zero percent rate. So that was the premise by which we were going with most of the definitions.

MR. DeBRANTES: So similar to your other measures.

MEMBER DELLINGER: Personally, empyema is one of the understood complications of pneumonia. It is the way in which some pneumonias even present. It is really -- and you know, we could go down to reach another

one. I am not even sure we had -- I guess we had similar complete lists of the PACs, but there was concern about that.

DR. RASTOGI: And the fact that you said it is a complication, that itself -- you know, you said it.

MR. DeBRANTES: So here is -- and I thought one that occurs naturally in the course of the disease. And I think this is the debate that we are having in the field, obviously, because we are instrumenting our program in quite a number of communities around the country, and it is something that we are having in the field and, I think, very useful and instructive to this committee; because instead of -- In fact, we don't -- There is no finger pointing or accusations or malfeasance or anything else in any of our implementations.

Instead, there is robust discussion within the physician and the hospital community around what truly can be

done to try to avoid some of these PACs, and I have difficulty -- and this is a difficulty inherent in any measure that relies solely on administrative claims data, is that sometimes you cannot parse the ones that are truly avoidable from the ones that aren't.

So our position has been, if the answer to the question is they can never be avoided, and I mean never -- so not one percent, two; they can never be avoided -- then we will remove those definitions from PACs.

If even one percent can be avoided, then we will include them, because if you don't, then you are sending a signal that says it is not important to count. We think that it is important to count. So that is the only point we are trying to make, and I understand the emotional and philosophical and other issues associated to it.

Our position has been let's count, and let's figure out collectively the extent

to which we can impact these numbers.

CO-CHAIR FLEISHER: Dianne.

MEMBER JEWELL: So perhaps this is just I need some clarification about the way things are indicated. I would agree that an empyema, by way of example, that occurs because a patient is out in the community with unattended pneumonia is a different thing than an empyema that occurs because of --

MEMBER TURNER: They didn't have the right antibiotic.

MEMBER JEWELL: Right. But what I also can appreciate is perhaps there is not, to use your word, parse. There is not a way to easily identify what presents on admission versus not.

So I am asking the question: Is that the problem or is it that, really in your mind, those things aren't different?

MR. DeBRANTES: I don't believe -- At least in our definition, that should not be the problem, because we are very clear that

elements that are present on admission would be excluded from counting as PACs. So that should not be a problem.

MEMBER PINDOLIA: Except it will, of course, because not all of us can tell you that was present on admission. But I am going to take another tack, and that is the public health perspective.

I would like to believe that we might be able to get people to come in so they aren't out in the community with an unattended pneumonia, and that I as a physician should take responsibility for that also.

Perhaps that is different being a primary care physician than some people who are, you know, a thoracic surgeon, but I think that that is a potentially avoidable complication, because that patient should have known to come in earlier. We should have made access available to them.

Now I can't change all of it, but there might be something I can do about it.

So I think we have to be very careful and not think too narrowly about health systems the way they currently exist, because they are really not very good. We all know that.

CO-CHAIR FLEISHER: Okay. Other comments? So, Patch, I am just curious. Is there any present on -- I mean, what I am hearing in Francois' response and your question, are there inclusion or exclusion criteria that could be applied, or certain CPT codes that maybe should not be in this measure that would satisfy some of your comments or concerns?

MEMBER DELLINGER: Well, I am not a CPT or ICD-9 expert. I couldn't possibly -- The numbers mean nothing to me. I need the labels or the descriptions. I understand what Francois is saying. I think that is reasonable. That was the biggest concern, I think, that the group had with this.

You know how to use this. It seems everyone will have PACs. All systems

will have PACs. I guess the issue becomes what is the range of PACs, and is there a way to change that.

CO-CHAIR FLEISHER: Lee?

MEMBER NEWCOMER: I think this measure I am going to compare to a coffee table book. It is designed to start a conversation rather than to actually weight or rank anything, because of all the issues that are described, you could get into that level of detail with every single PAC.

So if we think about it in that term, that this is simply a conversation starter and an internal measure for how we could get better or we could look at other systems and find that they have better measures, find out what they do to get ours better. But you can't make it perfect. It will not happen.

CO-CHAIR DUBOW: But I think -- I don't know if you are trying to narrow the use of this measure. I hope I am not hearing you

say that this is for QI and not for public reporting.

MEMBER NEWCOMER: Actually, that is probably exactly what I am saying.

CO-CHAIR DUBOW: Ah. Well, then I would disagree, because I think that this measure has, again as I observed about the other measure, I think this has salience for a patient.

I think that it may very well be -- -- I think that the value, actually, is that it is good for QI, and it does simulate improvement, because it will stimulate that conversation. But, by golly, this is very useful for a patient to be looking at performance, because it is absolutely understandable.

MEMBER HOPKINS: And to ask why?

CO-CHAIR DUBOW: And to ask why -- Well, to ask why or you say never mind.

MEMBER NEWCOMER: And to follow -- Can we follow that empyema example as maybe

why we are thinking differently about this?

I might be at Harbor General in L.A. where a lot of people simply don't have primary care and can't access and are not going to be on antibiotics, and get empyemas, or I could be at your general hospital in Nebraska where the doctors are using the wrong antibiotics, and empyemas are showing up.

Those are quite different scenarios, and I would not want to, as a consumer, assume that those two hospitals are exactly the same. They aren't. Same bag, but for much different reasons.

So it is a good conversation starter, but for us to say to a consumer hospital X is better than hospital Y based on these measures, I think, is a big stretch.

CO-CHAIR FLEISHER: That actually gets to how these measures are being utilized. Although I realize we all have that in the back of our mind and the doc well has emphasized that on several occasions, it still

-- Public reporting is different than pay for performance.

I know that may change. So the question is --

MEMBER NEWCOMER: But we are not talking about either one of those. We are talking about a consumer using the information.

CO-CHAIR DUBOW: But it is -- I think the difference is that what, obviously, I inferred correctly then, that -- but, unfortunately -- that Lee was suggesting that these are good quality improvement measures that should be used internally and not publicly reported.

Obviously, a criterion that we have at NQF is that these measures be used for both quality improvement and public reporting, and we have a disagreement. I believe that this meets that test.

MEMBER NEWCOMER: Maybe not. I don't mind them being publicly reported at

all, but I don't think they should be touted as a consumer measure that absolutely distinguishes a difference between one system and another. Whether they are public or not doesn't bother me one bit. It is a coffee table conversation started.

CO-CHAIR FLEISHER: So do you have any comment -- Well, go ahead, Larry.

MEMBER BECKER: So I absolutely think they ought to be out there in the public domain. They ought to be out there. I mean, there's any number of measures from cardiac measures to hospital mortality that make differences in systems, and people make real decisions about those.

So if there are potentially avoidable complications that one hospital more of those than another, then consumers ought to know that, and they ought to be making their decisions about that.

It may be with the specificity anybody would like, because it is not about

their specific decision, but we know it is about systems care. it is not necessarily about individual things.

So putting this information and having systems react to that and get themselves better because they are motivated, because the data is public, is hugely important.

CO-CHAIR FLEISHER: Helen, do you want to comment at all on this discussion? I am going to put you on the spot, Helen.

DR. BURSTIN: What do you want me to say? There is nothing else to say. I mean, I think the issue is just that NQF endorsed measures are intended for both, potentially QI as well as public reporting. It just is what it is.

CO-CHAIR FLEISHER: Barbara?

MEMBER YAWN: I just have a question about -- You said this was for people whose primary diagnosis was pneumonia.

MR. DeBRANTES: Discharge.

MEMBER YAWN: I know. Yes, the primary discharge. That is what I am going to get to. We all know that DRG, you rearrange the diagnoses, so you get paid the most. Does this remove some very important pneumonias from this, and is there any way -- and I don't know that there is an easy way to say, well, we don't want pneumonias acquired in the hospital; we want one that they had when they got in. Have you thought that through? You probably have.

MR. DeBRANTES: Yes. It is people who -- This excludes patients who got pneumonia while in the hospital. It is for people who -- if you get pneumonia during the hospital, you actually use a PAC.

MEMBER YAWN: Right. No, that is a hospital acquired, but there are people who come in primarily for pneumonia, but because you get a higher DRG, they are coded as something else as the primary diagnosis.

DR. RASTOGI: We don't want to use

the DRG codes in identifying our triggers. They are using the principal diagnosis code. So we deduct the DRGs from our trigger definition.

MR. DeBRANTES: Partially for that reason.

DR. RASTOGI: And we have given the triggers which are, but we then specified the AHRQ defined community acquired pneumonias, and if the principal diagnosis was that, that is what --

MEMBER YAWN: Okay. No, I wasn't worried about the in and out of the hospital so much as just rearranging an order.

CO-CHAIR FLEISHER: Other comments, new comments on new topics, because I would like to actually move to start voting on the criteria. No comments? Reva, take it away with a vote.

DR. WINKLER: So first going through the criteria for this measure, importance. It is a yes/no vote. So

important to measure and report. Everybody who say yes. I don't see any No votes. Okay, zero.

Scientific acceptability of this measure: Completely? Partially? Ten.

Minimally? There's got to be one. There you go. All right.

Usability: Completely? Seventeen?

CO-CHAIR FLEISHER: Hands up again.

DR. WINKLER: Okay, 16 that time.

Partially? This is usability, partially. Five.

Minimally? I'm still missing one.

Not at all? Okay. Seventeen for complete. Okay.

Feasibility: Completely? Eighteen.

Partially? Four. Okay, that should be it. All right.

CO-CHAIR FLEISHER: Okay. Next is

a vote --

DR. WINKLER: Recommend the measure.

CO-CHAIR FLEISHER: -- recommendation. Is there any conditions, just before? Okay. So either recommend Yes or No.

So, Yes? I get 20.

CO-CHAIR FLEISHER: No? Two.

Abstain? There should be none. Very good. Okay.

Next we have three from Brandeis, and this should be very similar to our discussions the last time while I was walking around San Juan. So I may have missed part. They didn't know that? I was walking around old San Juan. So, Reva?

MEMBER YAWN: Did you go in any of the breweries there?

CO-CHAIR FLEISHER: No. I was actually in Starbucks.

MEMBER YAWN: Okay. Too bad.

DR. WINKLER: Okay. Is Dr.

Tompkins with us? Oh, there you are. Sorry. Missed you. So we do have the developers. We have Chris on the phone with us.

Essentially, this is the same group of three measures that we have discussed for heart failure and we have discussed for AMI.

So measure 003 is the 30-day post-hospital pneumonia discharge ED visit measure. So that is the first one. The second one is measure 004, which is the 30-day post-pneumonia discharge E&M service visit measure. Then the last one, 005, is the 30-day pneumonia discharge care transition composite measure.

The methodology for these measures is the same as we have seen with the AMI and the heart failure. These are just applied to patients with -- correct me if I am wrong -- primary diagnosis of pneumonia at discharge. Correct?

DR. TOMPKINS: Correct.

CO-CHAIR FLEISHER: Well, once again we will get PACs here.

DR. TOMPKINS: Once again, just like with the VA, MI and the heart failure, this began with the same -- many of the same parameters as the existing CMS readmission measures.

So it has the same definition of what the cohort is, as currently seen in the mortality and the readmission measures. It has the same 30-day window, uses actually the CMS readmission rate measure as one of its components, and then asks the two additional questions: Were there an emergency department visit before any readmission, and was there an evaluation and management visit that occurred before either a readmission or an emergency department visit, if any occurred?

CO-CHAIR FLEISHER: Patch, any comments from the TAP?

MEMBER DELLINGER: I think the ID TAP was basically content. They had lots of

questions, and the only issue that really stood out was the rating on the composite measure, which is arbitrary, and is that the right weighting. There, of course, is not an answer to that, I guess.

CO-CHAIR FLEISHER: Any other comments? David?

MEMBER JOHNSON: The question is what are you comparing it to? Do you have a standardized index of people with the same disease that haven't been hospitalized? Is that your baseline comparison or -- Reporting a number could sound fairly onerous, but if you knew the likelihood of someone going to the ER with disease acts or comorbidities, what would be the -- Would that be the baseline comparison?

DR. TOMPKINS: Well, the general framework is to look at people who just came home out of the hospital. So the issue of people in the community who haven't been in the hospital is a separate reference

population.

The way these measures are constructed, it is using the nomenclature that is referred to as predicted over-expected. So there is a risk adjustment model that predicts the number of these events, of likelihood of these events, and then for the reporting purpose, for any given hospital, it is the extent to which it is predicted or, you might say, loosely speaking, observed rates differ from what is expected based on the risk adjusted value.

CO-CHAIR FLEISHER: David?

CO-CHAIR DUBOW: I just said we had been through all of that.

MEMBER HOPKINS: But since we are all together, I would like to put two pictures on the screen, page 67 -- Okay? This is the unadjusted data for pneumonia readmission rates: Very wide distribution, huge variability. Right?

Now please go to page 68 and see

what the risk adjustment algorithm has done to that distribution, and in so doing it has eliminated at least 50 percent of the outliers.

Now, you know, some will argue that, well, they should never have been identified as outliers in the first place, because it didn't take into account the size of the hospital and so on and so forth. But, really, how is the public being served by this kind of scrubbing of the original data so that in the end little or no variation remains? That is my question.

CO-CHAIR FLEISHER: Chris, do you want to comment?

DR. TOMPKINS: Well, I think that a few weeks ago it was when we were talking about this, right? I think that, to some degree, part of your business was you said that you were going to formulate either the Steering Committee, NQF or somebody who was going to communicate back to CMS about the

advisability of using this particular technique for outcome measures, and that is separate from whatever I did. Right? Our measure borrowed the existing thing. Philosophically, there -- The problem is that outcome measures have a lot of noise, and systematically, in general, there are three ways in which you can try to deal with that. One is to try to increase the amount of time, the number of events. You can increase the number of measures, which is my preferred philosophy.

This one, unfortunately, blends hospital-specific information with the growing mean, which is what a lot of people don't like. It results in discounting or granting a lot of -- or giving a lot of deference to the outlier status that is assumed to be noise.

MEMBER HOPKINS: And it is not the only part of such risk adjustment. That is the point.

DR. TOMPKINS: Right, and just for people who are following it, the risk adjustment, as it was occurring earlier and today, is typically thought of as the ability to use information such as the patient's comorbidities and so forth, to set an expectation.

This is rolled into what they refer to as a risk standardized method, which is hierarchical modeling that combines the individual and the group averages together.

CO-CHAIR FLEISHER: So this -- Would you like to comment from CMS?

DR. HAN: Is this the readmission, not the one that --

DR. BURSTIN: Yes.

CO-CHAIR FLEISHER: Just the readmission rates.

DR. HAN: CMS raised did raise this issue when the measure was developed. So this is what we understand, the narrow distribution after you risk adjusted.

First, we understand from the developer, the readmission rate is very high. It is like one out five. I am talking about a general AMI, heart failure, pneumonia. So it is like -- It is a bad thing across the board. It is really bad to have one out of five. So then we thought, that's fine.

The other thing is that we understand also from the developer, risk adjustment for comparing hospitals or profiling hospitals, there are certain factors we don't risk adjust them away. So this risk adjustment model, we have case mix in the risk adjustment model.

So what we were told is that readmission is very particular. Maybe the system factors play bigger role in the variation of the hospital performance on readmissions.

So that is the reason why that we got everybody is bad. So that is quite narrow, very close to each other, and system

factors play a bigger role. So that is why we were told that you can see the R-square is very low, because, you know, we purposely not to risk adjust system factors. That may tell you the variation.

I am not sure that I explained it well, because this is what we understand.

MEMBER HOPKINS: I think, actually, she probably right, but the last measure we just approved was not giving people a bye for system factors.

CO-CHAIR FLEISHER: So, Lein, will you comment, because we could spend probably the next two days debating this issue. So one option is to continue debating it. The other option is to call for a vote or a consensus that we ask NQF to continue evaluating the appropriateness of hierarchical models, but we defer to NQF to bring this up in other panels, and then --

MEMBER NEWCOMER: So my only new comment is that, going to the second

methodology makes the nation assume that one out of six is a normal, because that is what you have done. You have tightened that around and said one out of six is normal, but if you look at the preceding draft best practice, seven percent of hospitals are able to get to zero.

My point is that we should not accept the .15 as normal. We should be driving to zero.

CO-CHAIR FLEISHER: And to be honest, if you look at all the data on central line infections, we no longer accept what was the average.

MEMBER NEWCOMER: But we are encouraging that in this.

CO-CHAIR FLEISHER: So the question is, are you saying we shouldn't endorse this measure? This measure is actually already endorsed. So can I ask -- Amy?

MEMBER ROSEN: I want to get back

to the ER measure.

CO-CHAIR FLEISHER: Well, I just want to -- Does anybody have any other comments with regard to risk adjustment in the model?

MEMBER AMARASINGHAM: Well, I think the comment that is probably important is we have accepted this risk adjustment methodology for the other measure, the other composite measure.

So I think it is important for all the decisions we make that we are consistent, and if we have already set a precedent, I think I would be hard pressed to kind of go against that on almost an equivalent measure. But I do think that we should bear in mind how we voted on the prior measure, which is we accepted the composite but not the standalone measures.

CO-CHAIR FLEISHER: Okay. So I am just asking. Is there general consensus that -- Do you want to ask NQF to reevaluate the

hierarchical model in light of this discussion?

MEMBER HOPKINS: Yes.

CO-CHAIR FLEISHER: How many people think that NQF -- just a statement?

CO-CHAIR DUBOW: Well, it is not exactly NQF. It is CMS.

CO-CHAIR FLEISHER: It is CMS.

CO-CHAIR DUBOW: Isn't it?

DR. BURSTIN: Well, I think this is a bigger issue than we are going to be able to sort of swallow today. I think we have heard it. The committee has clearly indicated. I think those tables are incredibly compelling, in a way, but I have not seen it before, David. So I do think this is something we need to think about.

For now, though, that is the current endorsed measure that is not up for decision making today. It is part of a composite, just like it was on the first two.

CO-CHAIR FLEISHER: Thank you. It

sounds like in the report there will be a statement with regard to our concerns about risk adjustment. So new topics?

MEMBER AMARASINGHAM: Well, actually -- I think Amy was first. Go ahead.

CO-CHAIR FLEISHER: Sorry, Amy. New topic.

MEMBER ROSEN: Conceptually, I have some concerns about readmission as an outcome measure, and I think it is an outcome measure, and I think that being admitted to an ER, you know, within 30 days after hospitalization is as much dependent on the outpatient care one gets as the hospital care, and I don't see that taken into account here.

I think, conceptually, an ED measure is very different than a readmission measure. I think oftentimes it has to do with an availability of primary care and accessibility.

So I don't know how that is taken into account in this particular measure. So

I raise that concern.

CO-CHAIR FLEISHER: So, Patch, I will just take chair prerogative. When we vote, we will actually vote similar to our previous time where it could be endorsed only part of the composite, just to --

MEMBER DELLINGER: -- because there is a composite measure which measures E&M and gives you credit for E&M and deficit for an ED visit.

CO-CHAIR FLEISHER: Right. So it is identical. Chris, do you want to comment?

DR. TOMPKINS: Well, I think that the last comment about the composite is correct. This is a care transitions measure. It is saying when people leave the hospital. I don't think anyone is going to say that, when somebody left the hospital, that the hospital is solely responsible for everything that happens. The idea is to say that the hospital is part of the system.

CO-CHAIR DUBOW: This is a care

coordination measure.

CO-CHAIR FLEISHER: Right. So we will have the option to endorse each measure separately or endorse it only in the context of care or not endorse it -- in concert with a composite -- excuse me -- or not endorse it.

MEMBER AMARASINGHAM: I think what I would like to do for the committee is just restate our case for why we thought it needed to be considered together, and that is because there may be cases where a hospital does quite well on the readmissions, but maybe the ED measure or the post-acute care measure doesn't do so well.

I mean, you need to consider all of that, because there's very innovative models out there that I am familiar with, and that was the rationale for the composite, is for us to stem the measures previously, and I am not sure we would want to have different sets of criteria for the different measures.

CO-CHAIR FLEISHER: Thank you.

Any new comments? Okay, time to vote on the four criteria. Reva?

DR. WINKLER: Okay. We are going to start with the ED visit measure, which is 003, and so, as before, importance to measure and report. It is a yes/no vote. So all agree it is yes, ED visit measure --

CO-CHAIR FLEISHER: Just a clarification. So if we vote no, can we put it in the composite?

MEMBER NEWCOMER: Can you do two yeses? Can you vote for it and put it in the composite?

DR. BURSTIN: Yes.

DR. WINKLER: Right. But right now we are just doing the criteria and how this particular measure individually addresses the --

DR. BURSTIN: This is a standalone vote.

DR. WINKLER: Yes -- addresses the criteria. then the recommendation will have

those multiple components.

MEMBER JUSTER: Not just as a standalone vote. We do -- Should we just be voting no right now?

DR. WINKLER: The integrity of the measure itself, not how well you think it is going to work as a standalone measure. Right.

CO-CHAIR FLEISHER: So if you believe it should be in the composite, you should vote yes for importance. Is that correct?

DR. BURSTIN: You should vote whatever you think. I think the reality is just rate this measure as it stands by criteria. You will have the opportunity to make the decision of whether overall you want it as a standalone or -- you will get the chance to talk about it in the composite shortly.

CO-CHAIR DUBOW: We are voting on this measure.

DR. WINKLER: Right. So the

importance to measure and report for a measure of ED visits within 30 days after hospital discharge for pneumonia. Important to measure and report, Yes? I get 19.

No? Three. Oh, I can't add. I'm sorry.

Okay, scientific acceptability of this measure as specified. Does it meet the criteria completely? I see zero.

Partially? I don't see any others. So, okay, all 22 for partially.

Usability: Does it meet the criteria completely, how many? Five? Okay.

Partially meet the usability criteria? Fourteen.

And minimally meet the criteria? One. Is that right? No, I've only got 20. There is a No. Did everybody vote? You voted no? Okay, so you are a No. Does it still add up?

MS. BOSSLEY: That is right.

DR. WINKLER: All right, the last

one is feasibility. Completely? I get 20.

MS. BOSSLEY: And 21.

DR. WINKLER: Okay, 21.

MEMBER HOPKINS: Is this measure for fee for service, Medicare only? Is it one of those?

CO-CHAIR FLEISHER: Chris?

DR. TOMPKINS: The empirical estimations that we gave you are for fee for service only. There is no reason conceptually why it couldn't be used by anybody who is a payer or even a cafeteria delivery system. If you had information on all the covered services, you could implement it.

DR. WINKLER: Okay.

CO-CHAIR DUBOW: We only had 21.

DR. WINKLER: I was just going to say, feasibility: Did somebody have a partial? Okay, there is one.

CO-CHAIR FLEISHER: We are missing one usability.

DR. BURSTIN: In case everybody is

wondering, we have finally, in fact, ordered those little handheld voting devices. They are on their way, finally. We will see if that is better or worse.

CO-CHAIR FLEISHER: So now vote on the measure. Do you want to standalone for each of them?

DR. WINKLER: We have only just done the ED visit measure. Right.

CO-CHAIR FLEISHER: So go through the other two.

DR. WINKLER: So do you want to do the recommendations for this particular measure?

CO-CHAIR FLEISHER: Yes. So recommendations. The options are: Yes as a standalone; yes, but only as part of a composite measure; or no.

DR. WINKLER: Only as a standalone.

CO-CHAIR FLEISHER: Okay, just up or down. Yes as a standalone? Who votes yes?

DR. WINKLER: Eight. No for a standalone measure? Thirteen is what I get.

MEMBER HOPKINS: I didn't vote. I didn't understand what measure. The one on the screen is --

DR. WINKLER: This is the ED visit measure. You do? Okay, so that is nine. Nine yes, 13 no as a standalone measure.

CO-CHAIR FLEISHER: Okay, next.

DR. WINKLER: So next we will go to the pneumonia, which is the E&M visit measure. Okay? You know the discussion.

Is there any discussion about the E&M measure before voting on the criteria? Okay. So I take it you want to do -- all right.

So importance to measure and report on a measure of follow-up care afterward, yes or no. Yes? Twenty-one, and Barbara is not here. So 21, okay. That makes the No zero.

Okay. Scientific acceptability of

the measure properties: Completely meets the criteria, how many? Zero.

Partially meets the criteria? Twenty-one, okay.

Usability: completely meets the criteria? One.

Partially meet the criteria? Eighteen.

Minimally meets the criteria? There is one.

And not at all meet the criteria? Is that you?

MEMBER HOPKINS: I have looked in this section under this measure. It is old 409, right?

CO-CHAIR FLEISHER: So you are abstaining or not?

MEMBER HOPKINS: I am happy to abstain.

DR. WINKLER: Okay. Abstain. So feasibility: Completely meets the criteria? I get 14.

Partially meet the criteria?

Seven.

Minimally meet the criteria?

Barbara is out. Okay. Great. That's it.

All right. So the last one is the recommendation as a standalone measure. Yes?

DR. TOMPKINS: Just to connect a couple of dots. The mediocre scoring on -- The partial on scientific acceptability and the sort of mediocre on usability -- is that wound up in this question about the hierarchical modeling? That is what I thought. I just didn't want go unspoken about that, that it was something else that was major going on.

DR. WINKLER: Okay. So we are back. We are voting on recommendation of this measure as a standalone measure. We will get to the composite.

All in favor of it as a standalone measure? This is the E&M visit after pneumonia discharge. Four. I am getting four

yeses.

CO-CHAIR FLEISHER: And No?

DR. WINKLER: And No: Seventeen.

CO-CHAIR FLEISHER: Barbara is not here.

CO-CHAIR DUBOW: So that is correct.

DR. WINKLER: Fine, so that is everybody.

All right. So the next discussion is around the composite measure, which brings together the currently endorsed readmission measure with the ED visit measure and the E&M service measure, with the same weightings as we saw previously in the other two measures. It is a -4 for readmission, -2 for ED, and a +1 for the E&M service. So it is all the same, no changes.

CO-CHAIR FLEISHER: Did you want to have a separate discussion? I had thought we -- Any new points? Okay.

So importance: Yes?

DR. WINKLER: It is unanimous. So that is 21. Barbara is still out. Right?

Okay. So scientific acceptability: Completely meets the criteria? Two.

Partially meets the criteria? Nineteen. Okay.

All right. Usability: Completely meets the criteria? I am seeing none.

Partially meets the criteria? Is that everybody? Okay. So it is 21.

Feasibility: Completely? Thirteen.

Partially? Eight. That's it.

Okay, so the next is the recommendation on the composite measure going forward. Are there any conditions?

CO-CHAIR FLEISHER: None.

DR. WINKLER: Bless you.

MEMBER HERMAN: We did talk about a condition when these went through the TAP, is that should this always be tied to an E&M?

We talked about the carriers, and there is a lot of places that are doing this through home visits and things that are not tied to an E&M.

DR. WINKLER: And for those of you who put that in your comments, you got a No on the measure and the composite.

MEMBER HERMAN: Right, but --

DR. WINKLER: But it still passed.

MEMBER HERMAN: Yes.

DR. BURSTIN: Although I was actually going to raise the same thing, not so much for the composite, but I do think it is important, particularly for Brandeis and CMS, to consider the fact that I think this measure would have done better as a standalone, if in fact that issue had been addressed.

I think that is partially the reason why it had difficulty, because it excluded all the innovation of people, in fact, using a nurse telephone or something like that, when you wouldn't capture it as a physician E&M visit.

MEMBER JOHNSON: Is there any way that you co-adjust for utilization of ERs in a given area, because if certain areas -- and this came up before in the discussion of coverage -- it is a standard. Somebody calls after the office is closed, go to the ER, as opposed to management interventions that may preclude going to the ER, and whether or not that ER actually led to a hospitalization.

The utilization of the ER -- is that variance part of this? How do you balance that as far as what standards are for people that call after hours?

DR. TOMPKINS: Well, I mean, we would lose a lot of points on feasibility if we tried to capture things like telephone calls and information about whether the primary office, doctor's office, was closed when the ER visit occurred.

So there is always a trade-off here. This -- as I said before, it is a profound system, and where the system breaks

down is going to ding you somehow, and if the primary care doctor's offices tend to be closed and the ED is the only thing that is open, then that system would tend to look worse because of that.

If the system wants to improve in some way, it ought to do its own engineering to figure out why it is that it is either below average or less than it thinks it ought to be. So issues like access to alternatives, I think, is part of the game here, and gain. That is the opportunity for gaming.

MEMBER JOHNSON: But I guess, if it is a system that is unified as opposed to a hospital when you've got independent providers. Then it is not really a system. It is just a -- It is an organization of dysfunctional participants.

DR. TOMPKINS: Well, you know, we've turned a corner here, and ACO is the great acronym of the day, and people are asking how is it that we can have a fragmented

system look more like a system.

Measures like this are intended to profile the system performance, and leave it to the professionals to figure out where the deficiencies are that could best make their score improve.

CO-CHAIR FLEISHER: So calling for a vote. All those in favor of the measure -- The composite?

DR. WINKLER: It is unanimous.

CO-CHAIR FLEISHER: Okay. I guess we are now going to call for any public comment on the measures that we have discussed today, since this morning.

DR. GALLAGHER: This is Rita Gallagher from the American Nurses Association.

There is no specific comment on the measure as discussed, but really a reminder that consideration of patient outcomes in the absence of consideration of the processes and/or structures by which those

outcomes arose is problematic.

CO-CHAIR DUBOW: Thank you, Rita. Anybody else?

We are due for a 15-minute break. That means we have this afternoon before we eat dinner, we need to finish the cardiovascular and the surgery measures.

So you are getting 10 minutes. We are going to start again at four, and we are going to start with the BET measures to take advantage of the fact that we have the developer with us today.

(Whereupon, the foregoing matter went off the record at 3:50 p.m. and went back on the record at 4:05 p.m.)

CO-CHAIR DUBOW: We are eating at six o'clock, because nobody wanted hang time in the Marriott, and Heidi has, obviously, worked that for us. Then the bus will come and bring you back at -- We are leaving at 7:30, but maybe we will make it earlier.

Tomorrow morning we are starting

promptly at 8:30, and food will be here. So the buses will get you. You will be picked up at 8:10, and good luck to you in coming downtown at rush hour to be here. I said good luck.

This is the David Johnson dinner at six o'clock. Okay. Let's get started, guys.

CO-CHAIR FLEISHER: So we are going to start the cardiovascular measures, and we are going to start -- Amita, do you want to join us at the table. Why don't you start with the Bridges to Excellence measures, and then we will go to the STS and then the SVS measure.

Do we have the TAP Chair?

CO-CHAIR DUBOW: Dr. Gibbons.

CO-CHAIR FLEISHER: Dr. Gibbons, are you still on the line?

MEMBER GIBBONS: This is Ted Gibbons.

CO-CHAIR DUBOW: Are you there?

CO-CHAIR FLEISHER: Great.

MEMBER GIBBONS: I have been here all day.

CO-CHAIR FLEISHER: Fantastic. Dr. Gibbons, if you would like to vote on any of the measures, I guess you can send an email to Reva.

MEMBER GIBBONS: I will send an email. Fantastic.

CO-CHAIR DUBOW: Thank you.

CO-CHAIR FLEISHER: Greatly appreciated. So I guess, Reva, do you want to start with a brief introduction of the two measures, and then --

DR. WINKLER: All right. The first two measures we are going to talk about are very similar to ones we have already done. These are more of the Bridges to Excellence measures with PACs.

The first one is the proportion of patients hospitalized with AMI that have a potentially avoidable complication during the

index stay or in the 30-day post-discharge period.

So the methodology is the same. The approach is the same. The number is OT1-030-09.

CO-CHAIR FLEISHER: Ted, do you want to -- Any comments from the TAP?

MEMBER GIBBONS: Well, the TAP had differing approaches or at least different takes on the AMI or stroke avoidable complications.

The AMI had very few controversies -- There was some feedback that weren't controversial, but really reflects some of the questions that were posed earlier, primarily--

CO-CHAIR FLEISHER: We are losing about every other word.

MEMBER GIBBONS: I'm sorry. Can you hear me now?

CO-CHAIR FLEISHER: Yes.

MEMBER GIBBONS: I'm sorry. I am on a land line. It is actually connected with

wires and walls.

The stroke issues related to the understanding that some of the adverse outcomes of stroke may be occurring -- I think those group of issues have kind of been discussed earlier. I don't know that there were -- what are seen as complications of care are actually part of the -- process --

CO-CHAIR FLEISHER: Thank you. Amita, do you have any comments or responses to the concerns of the TAP?

DR. RASTOGI: No. Dr. Gibbons commented very well. Francois is not here right now, but what he mentioned to me was the same thing, comorbid considered as part of the care process of stroke and cancer, and I kind of basically thought that it is important to count it, even if it happens, because in the end we want to reduce the number of PACs. So it is a complication that could be part of more discussion.

MEMBER GIBBONS: Yes, but I think

that the discussion really homed on the global access to care and the early recognition of stroke, and trying to reduce the morbidity associated with stroke y focusing on the global process.

CO-CHAIR FLEISHER: So were you on the line or were you available to the TAP so that the comments after -- I was just wondering if the TAP got the responses from Bridges to Excellence and what the TAP thought of the responses.

MEMBER GIBBONS: Yes, the discussion that took place on the responses really focused on really some of the same issues, such that they weren't meant to penalize individuals institutions that took care of high risk stroke, but really to reflect the process of care and reduction of potentially avoidable complications.

CO-CHAIR FLEISHER: Comments from the Steering Committee?

MEMBER AMARASINGHAM: I just had a

question. I assume the risk adjustment methods are essentially the same as before, and I guess my same question that I had before, which is: These risk adjustment methods which appear to me novel have not been tested in head to head comparisons with other risk adjustment methods yet. Right? That is being done by Rand?

DR. RASTOGI: That is right, and as Francois mentioned, you know, the PAC rates are developed, and then the severity index is there. If folks feel strongly about it, they can adjust it for the severity index, but that is basically what it is.

Rand currently has taken our models and are giving it a complete make-over, so to say, testing the models and doing all kinds of analyses and subgroup analyses. So that is a different project completely.

MEMBER AMARASINGHAM: I guess the question would be, because it does appear to be sort of novel approach to do this, and I

recognize the challenges with data collection which requires this approach, what will we do if Rand concludes that it is inferior to other risk adjustment approaches?

DR. RASTOGI: At this point, I don't think it is a matter of whether it is inferior or superior. It is just finding a risk adjustment approach. Right? So at this point, the PAC percentages that we are calculating with the patients, that calculation is there. Whether they are severity adjusted depends on the users, whether they want to use it or not.

So the emphasis is more than counting the PAC rate. That is the idea behind it.

MEMBER AMARASINGHAM: The only thing that concerns me -- I love the approach, and I think what Prometheus is great, but the reason I had to vote no on the other cases is because there is an element of a leap of faith here about this risk adjustment method. It is

not a typical approach, and it hasn't been validated against other approaches.

MEMBER PINDOLIA: I wonder if the other approaches have been validated in this type of a measure, though. Yes, they may have been done for a long time, and most of us who use Charleson now, but it has been done for a long time, but it is really not a very great measure, and some of the others that have been used maybe don't work terribly well for this particular situation. So --

MEMBER AMARASINGHAM: Well, I agree. You know, with any measure you want some sort of reproducibility and triangular validation. So if you had multiple measures that are pointing to the same result, even if they are inferior but there are somewhat all pointing to the same result, then you feel like you are on stronger ground.

MEMBER PINDOLIA: I agree. Yes.

CO-CHAIR FLEISHER: Okay.

MEMBER TURNER: So one of the

axioms of the panels when we were reviewing the statistics in papers is how they developed the models.

I understand the p value driven models, and that is exactly what it looks like it is. You only include things that are in there with a p value that is less than .25, and then you use stepwise modeling, which is also a no-no, from our point of view, the reason being that you should have clinical input to try to decide which variable is important, not p values, and not let the program decide what variables to be in there.

So it makes me feel much more uncomfortable about the risk model if that is the way it was developed, because it doesn't really have a clinical mind behind it.

MEMBER HOPKINS: Do just a question of philosophy. Because clinicians think that the factor is important, is that sufficient, even though you put it to a statistical test and find that it isn't?

DR. TURNER: Yes, because there are a lot of factors that affect each other, even though it is not obvious by itself, that a p value is like, let's say, age may not be significant in the model, but it has an important affect modification on other things that you are concerned about, like the diseases.

So it in itself isn't independently predicted, but the way other important variables that are in the model are interpreted has to have age in the model to correctly understand them.

The other thing is that the power of this is that you have really big databases. I mean, you've got lots of ability to include a whole host of variables, and I think that is where you have the opportunity to really include a lot of the key predictors and not let the machine decide what is in there.

MEMBER HOPKINS: Can I ask Sean to comment on this, as the expert on risk

adjustment. I thought it was right to use statistics to demonstrate what is significant or not in a risk model.

DR. O'BRIEN: Yes, and I think you can find that clinicians do not always know exactly what is going to predict. There are all kinds of surprises, and the surprises aren't always just spurious associations. They are real things that are uncovered by the data that are not uncovered by clinicians.

On the other hand, depending on the size of the data, what outcome you are studying, the variable, selection procedures can be highly unreliable in the sense that if you did the exercise in choosing variables and you repeat it again in a different dataset that based on the same population, you may get entirely different predictors in the model.

I think it is acceptable, although based on a little bit of error. You are not going to choose the variables perfectly, but one consideration. As these models become

published and codified and accepted as the model. Other people will come along and want to do a different approach, and we say, no, we have endorsed this model, but there's variables in the model that are partly haphazard.

So you might think you don't know exactly what the important predictors are, because there is some uncertainty in the approach that basically takes the potential variables that you think are important and just leaves them in the model and then adjusts for them is another possible approach to doing it. So not relying so much on variable selection is an option that I think some statisticians would support.

CO-CHAIR FLEISHER: Amita, did you want to comment?

DR. RASTOGI: I would like to give a little bit of feedback on that, because as a -- you know, I am not a biostatistician, but as a physician, I was concerned about the

reliability of each variable that went in. And as you will see, each model that was developed -- we've got 21 UCLs now. Each model that was developed for each part of that portion or part of that episode was very carefully calculated for input variables that are specific to that particular episode.

So AMI variables are not the same as pneumonia variables, and they are not the same as stroke variables. There are certain -- and you will see that in the expanded trigger in the all-codes workbook. Each one of them was specifically tailored for that particular episode, from input from the physicians who helped us develop it.

MEMBER TURNER: No. I think that is great. What I have trouble with is when you take your risk factor variables that your experts give you and then you enter it into the model using stepwise regression, and drop them out based on a p value.

I thought also another thing about

what you do is you enter classes of variables into the model and see how they affect it, and I like that, too. I just feel like this is a very -- It is very important to not allow the modeling to be too specific to the data that you have, for the reasons we just heard.

CO-CHAIR FLEISHER: Barbara, and then I have a comment.

MEMBER YAWN: I would have been a lot more concerned if the p value that had made them remove it had been .05. When you start getting to .25, then I think that I am much less concerned.

I think then you may have the best of both worlds, which is trying to look at some ability to do more than use our opinions, which, of course, are wonderful, but we are somewhat narrow minded occasionally about, you know, I see family medicine, so cardiology must look just like family medicine or things like that.

So I was pleased that there was,

in my opinion, this combination of the clinical significance and then some very broad statistical significance. I mean .25 is really -- you know.

MEMBER TURNER: So to make all of us happy, did you do this in multiple datasets and find that the same variables were, in fact, selected?

DR. RASTOGI: There was a bootstrap validation process that was followed, and it was run 200 times, and then the variables that they selected 80 percent of the time were the ones that were finally chosen.

MEMBER TURNER: This is the same data, but that is good.

DR. RASTOGI: Yes. And then to your point, what we have now -- These programs are now available as an automated function, and each health plan can develop it from scratch on their own data.

So they don't have to use the

coefficients and the variables that we select, just the EEGs and all. When I was with Ingenix and I developed the EEG regression models, you publish the coefficients every two years, every three years. Right? Then everybody has to adopt it, or at 3M. You know, the publish the coefficients every so often, and everybody has to use that as the industry norm.

Here it gives the capability for each health plan to choose the variables that are selected for that population. So it runs from scratch, but they have to have a minimum sample size for that to go. So we give them scoring one, scoring two, scoring three logics, which is all part of the automated programs.

So from usability point of view, the needed criteria are friendly, and the HealthPartners has run it, and they have really -- You know, I was talking to Chad Hines just the other day, and they have really

enjoyed working with it.

CO-CHAIR FLEISHER: So, Amita, just so that we understand. So we are endorsing a measure using PACs over a time period. The risk adjustment seems to evolve over time, and that is -- or it can be -- In other words, getting back to your question, Ruben, if you go with Rand and find that the risk model should be tweaked, would you come back and actually say the risk model should be different, although the basic measure stays the same?

DR. RASTOGI: Reva, I will throw that question back to you, whatever you recommend.

DR. WINKLER: Essentially, whenever there are significant changes in a measure specifications, measure developers are expected -- it is too strong to say obligated, but to tell us that that is what they have done and that may, depending on what those changes are, prompt an ad hoc review. Is that

right, Helen?

DR. BURSTIN: That is correct.

CO-CHAIR FLEISHER: So it sounds like, if we endorse this measure, that if the risk model changes, then you would actually be obligated to come back.

MEMBER AMARASINGHAM: I think I would amend my word on the previous one.

CO-CHAIR FLEISHER: You can do that. Amy, comment?

MEMBER ROSEN: I want to make a comment on the risk adjustment methodology, in that typically we don't include treatments or -- I think it is listed up there as various types of services, in the risk adjustment model, because those are very subject to physician practices.

So that is just a thought I want to throw out there in terms of, if you are thinking about doing it, perhaps now including that in the risk adjustment model.

CO-CHAIR FLEISHER: Are those

measured during the episode of care, like before?

DR. RASTOGI: That is right, and some of the services that you mentioned like durable medical equipment and all were deliberately put in partly because they were used as surrogate measures for the patient's condition.

It is very hard to put in a diagnosis code to say that they are wheelchair bound, you know. So the DME was kind of a surrogate to suggest that the patient was a little more debilitated than another person.

So these were the type of procedure or CPT codes that were put in.

CO-CHAIR FLEISHER: You almost have an expert group who wants to help you with your risk model, or at least see the Rand paper when it is ready.

MEMBER ROSEN: One thing I want to comment on is we are focusing a lot on the risk adjustment, but I am concerned about the

PACs. Do you have -- What I am confused about is exactly how they get defined for each disease or each chronic condition and each different type of disease.

Are there different PACs for stroke, different PACs for AMI, and how does that get determined? Who determines that? Is there a clinical expert panel that meets? Is there some kind of Delphi process that happens or how does that work?

DR. RASTOGI: Yes. Some of it is detailed in the document which Francois kind of circulated called History of PAC Development. So there were working groups that were appointed for different ECIs which gave their input, and then in some of the ECIs where I got more input, I worked with the physicians one to one, back and forth, sometimes with a panel, sometimes with individual physicians, to get the right coding, circulated the documents to work with them to get the input and feedback. So that

is how these came to be.

MEMBER ROSEN: I am sorry that I missed that.

CO-CHAIR FLEISHER: Not a problem.

MEMBER ROSEN: It still makes me a bit worried thinking about it, because you can get a group of -- As you all know probably, you can get a group of physicians in a room, and you know, the next day another group of physicians will come out with exactly the opposite.

So to some extent, it is much better that we are able to see both clinical and empirical testing of outcomes that one develops. So that is in the ideal world.

CO-CHAIR FLEISHER: Thank you. Sean?

DR. O'BRIEN: At this time, I have got one comment for the PAC measures and one that is specific to AMI.

The one that applies to AMI is that looks like patients undergoing CABG are

excluded, and it seems that is a factor that is definitely under the control of the provider and may be why certain types of units or hospitals who promote CABG may have different outcomes. You may adjust that away. I guess I will stop there.

DR. RASTOGI: Yes. You may have episode, but it is a very short episode. It only has 30 days-plus and discharge period, and patients who had to have CABG -- we thought of them as slightly different than the rest of the population with AMI, something to the effect that mortalities are excluded, you know.

We have to know that up front going in, but we are looking at these PACs only at patients with AMI who didn't have a surgical intervention.

CO-CHAIR FLEISHER: Any other comment?

DR. O'BRIEN: The second comment is that there is a component endpoint which is

useful, but if you look at that, the later statement in the submission basically says it is a percent of patients. You capture patients that have a PAC.

Then if you really want to un-PAC it and know what is considered a PAC, you have to go to the Excel file and look through all these codes. There is not really any English language, concise statement of what are all the PACs that go into the PAC, and I think for interpretation and reporting and transparency, it is important to have kind of a more understandable statement of what you are measuring.

DR. RASTOGI: That is exactly right, and I think Francois and Joyce already talked, and they are going to write a one-pager for each of these measures, so that for public comment it becomes a little bit easier to understand.

CO-CHAIR DUBOW: A crib sheet version, I would guess.

CO-CHAIR FLEISHER: Yes. Any other comments? Ted, do you have any other comments, having heard this discussion, from the perspective of the Steering Committee --

MEMBER GIBBONS: I just want to comment that I think it points up the fact that we need to keep reexamining how the risk adjustment might change over time, but I think that many of the --

CO-CHAIR FLEISHER: We are losing you again.

MEMBER GIBBONS: I'm sorry. Many of these comments really reflect the fact that we have to keep track of how it evolves over time. I haven't heard any comments that depart significantly from the TAP discussion.

CO-CHAIR FLEISHER: Yes, thank you. So are we ready to vote? Any other comments? Any public comment? Hearing none--

DR. WINKLER: Okay, we will vote the measures independently, so separately the AMI, and then we will follow with the strokes.

So the first one we are going to talk about is the AMI measure. So everybody got that clear? So the first criterion is the importance to measure and report. It is a yes/no. How many say Yes? It is unanimous. Okay. So that is 22.

All right, scientific acceptability of the measure properties: Completely meet criteria, how many? All those for completely? Ten.

Partial? Twelve. Okay. That is everybody.

All right, usability: Completely? Completely for usability: Ten.

Partially? So everybody else? Yes, it is everybody else.

All right, feasibility: Completely? Fifteen.

Partially? Seven. that's it.

MEMBER JOHNSON: Never let it be said that you don't get your exercise.

DR. WINKLER: It's lovely. Thank

you so much. So recommendation: Are there any conditions?

CO-CHAIR FLEISHER: First, anybody want to propose any conditions on the recommendation? No. Okay, so it is a simple yes or no vote. Those voting Aye?

DR. WINKLER: Twenty-one.

CO-CHAIR FLEISHER: No?

DR. WINKLER: Okay. One No. That means no abstentions. Okay.

So now we move on to the stroke measure, the same measure with the PACs for stroke.

Importance to measure and report: Yes? I am getting 19. So it is 19.

No? You are abstaining. Barbara is gone. Okay, so you are going to 20, and there is one abstention, and Barbara is out of the room. Okay.

CO-CHAIR FLEISHER: Can you turn the microphone off for Barbara?

DR. WINKLER: Okay. So for

scientific acceptability of the measure properties, how many think it completely meets the criteria? This is the stroke PAC measure, completely meets criteria. Four.

Partially? Seventeen, and that is everybody.

Usability: Completely? Four.

Partially: Seventeen. Yes, it is everybody else.

Feasibility: Completely meets? Fourteen.

Partially? Seven. That is everybody.

So now the recommendation.

CO-CHAIR FLEISHER: Any conditions? No. Okay. Those voting Yes?

DR. WINKLER: Yes for the measure? Okay, so it is 20 Yes. Amy, your vote is a No? Okay.

CO-CHAIR FLEISHER: So next could we have the STS measure? I know we have Bruce on the line.

DR. WINKLER: Bruce isn't from SVS.

CO-CHAIR FLEISHER: No, Bruce -- He is next, ACS.

DR. WINKLER: Were we expecting anybody?

CO-CHAIR FLEISHER: But Sean should be able to -- Are you the methodologist on the STS measures?

MEMBER GIBBONS: Yes.

CO-CHAIR FLEISHER: Okay.

CO-CHAIR DUBOW: Do we have anyone from STS on the phone?

MS. HAN: This is Jane Han.

DR. WINKLER: Do we have the measure up? This is a composite score of endorsed measures for coronary artery bypass from STS. This measure is a combination of outcome and process measures that have been endorsed by NQF for a while, several years, actually.

So this measure presents a

composite methodology for combining these measures into a single score. And that is kind of what it is.

CO-CHAIR FLEISHER: That is it. Do we have any comments from the TAP?

DR. WINKLER: They liked it.

CO-CHAIR FLEISHER: They liked it. Okay. Sean, any comments? Any questions?

DR. WINKLER: One thing, I asked Amy to take a look at this measure as a methodologist to see if she had any comments about the composite methodology.

CO-CHAIR FLEISHER: Just tell us if you do vote no.

MEMBER ROSEN: So this -- It is kind of interesting after sitting here listening to some of the comments. But this measure -- The composite includes both process and outcome measures. So I want people to be aware of that.

So we have -- So that is okay, but just so that you know, we have the mortality

and morbidity measures as well as the perioperative process of care measures and the operative care measure. So those are all included in the composite. The composite is -- It is fine with me, but I just wanted to point that out.

The composite is not weighted, which is different from the other composites we have seen, and again I kind of like that myself, because I think that the weighting schemes that we have seen here are really just clinically based and isn't evidence of empirical testing.

So as far as I am concerned, weighting them equally is good, and I think it is important to -- in thinking about a composite measure, to think multi-dimensionally, and I think looking both at processes of care in relation to CABG surgery is quite relevant.

What that does, however, on the other side of it is that we lose focus on what

the specific outcome measures might be, because we are looking multi-dimensionally.

So overall, I thought that there was really adequate empirical testing as well as clinical oversight, and that the developers did a very nice job in putting the composite measure together.

So maybe I should go to my room now.

DR. SHAHIAN: This is Dave Shahian from STS. Could I respond to the comments?

CO-CHAIR FLEISHER: Sure.

DR. SHAHIAN: Just wanted to point out that, although we do roll up those process and outcome measures to a single score, that when we present the scores, they are presented individually as well, and the providers are also given a granular view of where they failed specifically in the all or none process measures to help them to inform their performance improvement activities.

So we provide both the overall

score, but also a very detailed drill-down to the component level.

MEMBER ROSEN: That is absolutely correct, and I guess the thing that is somewhat puzzling is this is used at the provider level as well as the practice group level and at the hospital level, and I don't know to what extent comparing providers has been tested statistically in terms of what panel size providers would need.

It is very different to compare providers versus hospital level measures.

DR. SHAHIAN: Well, let me just comment on that. We base these scores on what we call an STS participant. Now in the vast majority of cases, an STS participant is a hospital with a cardiac surgery program.

There are, however, instances in which there may be, for example, two large cardiovascular surgery groups within a hospital that may decide to independently contract with STS to have their results

measured and computed using the STS methodology. But the unit of analysis is an STS participant.

CO-CHAIR FLEISHER: Okay. Joyce had a comment?

CO-CHAIR DUBOW: Well, I have a question. That is interesting that it is an STS participant. It means that that person has to participate in the registry.

DR. SHAHIAN: That is correct.

CO-CHAIR DUBOW: There is a fee?

DR. SHAHIAN: That is correct.

CO-CHAIR DUBOW: Isn't that correct?

DR. SHAHIAN: That is correct.

CO-CHAIR DUBOW: But that wasn't my point.

CO-CHAIR FLEISHER: The percent of people who participate in the registry is --

DR. BURSTIN: Over 90 percent.

CO-CHAIR FLEISHER: So 90 percent of all CABGs, I think, are in the registry.

Correct? Sean?

CO-CHAIR DUBOW: Okay. But my question has to do with my understanding that the reporting of this measure is wrapped up in the measure that we would be endorsing or recommending for endorsement, because -- and this is a question. I am just trying to get it out.

That is unusual. We don't normally -- This has the star. If you read the presentation, this has the three stars with 77 percent of physicians or STS participants falling into the two-star category, as I recall.

So that essentially this measure, as I understand it, encompasses the way this stuff is reported, the results are reported, as well. Is that correct?

DR. SHAHIAN: Well, do you want me to respond to that?

CO-CHAIR DUBOW: Please.

DR. SHAHIAN: That is the way we

submitted it, although each provider is also given a numerical score, both overall and for each of the four domains of the composite. So one could use that numerical score.

Most of our providers have found it useful to also have this star rating, which we have developed and which has shown itself to be fairly consistent among providers over time, and also to correlate fairly well with the performance in each of the individual domains.

So, yes. I guess the answer to your question is, yes, we did submit this along with the star rating, because that is the way it has been operationalized.

CO-CHAIR DUBOW: Okay. So as I recall, the reporting is at the 99 confidence interval, and I think that is what I remember reading. But my question is whether the ratings themselves are publicly available or is it just the stars?

In other words, if a reporter

wanted to take the numeric values, would those be available?

DR. SHAHIAN: I don't know whether that has been in our discussions or if it was Consumer Union and for our own internal reporting.

One of the concerns we have, and I have seen this, is that external entities sometimes take numerical scores and do funny things with them that they were never intended to do. They may take, for example, the patients that we categorize into three groups and they may try to change them into some other weighting system that may not be appropriate.

So I would have a little bit of concern about that. You know, we've spent an awful lot of time doing pilot studies using this three-star system. We have been using it in practice now for over three years. We know how it works, and we think it is sound, both theoretically and in practice.

I guess my concern would be to put numbers out there which somebody could irresponsibly and incorrectly use.

CO-CHAIR DUBOW: Not to beat a dead horse, I think we ought to consider -- You know, I think this is -- I would ask the staff if this is not unusual, that we buy into a reporting approach in addition to the measure itself. I don't remember another measure like that. So I wonder whether we ought to separate them or whether we can separate them. I don't know.

MEMBER AMARASINGHAM: Well, I am just curious whether it is like some of the Medicare measures that we have debated, that they were validated deriving data on the Medicare population. We are saying that this was sort of derived and validated on STS.

CO-CHAIR DUBOW: We don't tell Medicare how many stars to attach to a hospital.

CO-CHAIR FLEISHER: So I would ask

David, so if this committee based -- One of the potential votes is to endorse this on the condition that the star rating not be part of it. If that condition is required, would the measure developer accept that?

DR. SHAHIAN: Well, I would think we would want to seriously think about that, and I would not want to make that decision right now, because as I said previously, we would be concerned about misuse of the raw numerical data. So I think that is something we would want to take under consideration.

CO-CHAIR FLEISHER: Okay. Well, thank you. We will take that into consideration as we go.

DR. SHAHIAN: Sean O'Brien who worked with us and was very instrumental in helping to develop it -- I realize that he is an awkward position there today, but perhaps for informational purposes you might ask his opinion about the question we have been discussing.

CO-CHAIR FLEISHER: So as your are squirming in your seat, Sean, do you have a comment?

DR. O'BRIEN: Personally, I think NQF should go in the direction of endorsing -- All the aspects of reliability and validity and usefulness can only be assessed in a particular context. A measure has to be used for a particular purpose. It needs to be particular.

I think that, even if you want to basically have a more broad description of what you are endorsing, demonstrating that it is useful in at least one particular application is helpful.

So I think you can either take this as an example of the usefulness of the measure or you could decide to actually make that example part of what is being endorsed, either way. But I think it is a demonstration of how it has been used in practice, and it has actually been well accepted and used by a

lot of groups.

CO-CHAIR FLEISHER: Anne?

MEMBER DEUTSCH: I just want to add that I actually like that they put effort into making it more understandable potentially to consumers. It helps us interpret the information perhaps a little bit more, and so what is an important difference.

One of the projects I do, we show some people in senior centers some quality measures that Medicare has put out, and we ask them to interpret it, and some people say, oh, only five percent difference, that is nothing. Other people say, oh, five percent difference, that is important.

So this, I think, actually helps potentially people to interpret the data and what is an important difference.

CO-CHAIR FLEISHER: Okay. New comments? David?

MEMBER HOPKINS: I have some questions and some comments.

So first question is for NQF: All four of these elements, the individual measures within the composite, have been endorsed. Is that correct?

DR. BURSTIN: Correct.

MEMBER HOPKINS: Okay. Second question for the developer: Did I understand that the weighting of these four elements is equal, is 25 percentage?

DR. SHAHIAN: Well, we did not weight them, but by virtue of the variation in each one of the individual domains, they were standardized, and by virtue of that the mortality measure does end up carrying more weight within the composite. But we did not start out by saying we want more mortality to be weighted more heavily. It is purely a function of the standard deviations.

MEMBER HOPKINS: Okay. So what are the weights for these?

DR. SHAHIAN: Sean, do you want to comment on that?

DR. O'BRIEN: We first standardized each domain's score at a common standard deviation, and then we applied equal weights to those standardized scores. So in that respect, the weights are one, one, one, one, actually one-fourth, one-fourth, one-fourth, and one-fourth.

MEMBER HOPKINS: Okay.

DR. O'BRIEN: Now it is also true that there is no such thing as no weight. Any weighting system, even if it is equal weighting, has implications, and I think we strived to make the implications by weights -- to understand them internally, and to publish them and basically make them apparent.

Part of some of the implications, as basically he is explaining, is that on the original raw scale of mortality you have equal weighting on the standardized scale, but that implied unequal weighting on some other scale. It is not possible to have equal weighting on every single scale. If you have equal

weighting on one scale, you have unequal weighting once you standardize them. So it is a situation that has no solution.

So what we did is we basically described the implications, and one percentage point difference improvement on the mortality scale for the mortality domain would essentially increase a provider's score on the composite about the same amount as an eight percentage point increase in the morbidity domain.

MEMBER HOPKINS: And how about the use of IMA?

DR. O'BRIEN: Did you have the numbers in front of you? I think it is somewhere around 15 percent.

MEMBER DELLINGER: So what you are saying is a difference of one is one standard deviation for each of the elements?

DR. O'BRIEN: Well, I was saying a one percentage point -- the difference between a two percent mortality rate and a three

percent morality rate is going to have a certain impact on your measure performance.

MEMBER DELLINGER: But you rate one standard deviation away from the median as the same in each of the four areas. Is that what you are saying?

DR. O'BRIEN: Right. Well, we took -- Each member has -- You calculate a score, and we rescaled each score at a common standard deviation. The way you do that is you divide each score by its standard deviation.

MEMBER BECKER: Could I ask -- Is that methodology you just described transparent and available to everybody?

DR. O'BRIEN: Yes.

DR. SHAHIAN: This was published in detail three years ago. I think you have the PDF of that.

CO-CHAIR FLEISHER: Sure. We want to keep moving. Go ahead.

MEMBER HOPKINS: So, I'm sorry. I

have a printed version. I don't know where it is in this package. Page 14 of something shows a really interesting analysis of -- and this is related to star ratings, actually -- of the stability of the star ratings.

I look at that, and I say only half remained where they were. I appreciate the fact that they didn't go down two levels or up two levels. Page 79 in this one? Thank you. But they changed by one, half of them.

Considering the fact that this is built on 99 percent or 98 percent confidence, to me, that suggests that there is something screwy about weighting these particular measures altogether.

I don't know how else to interpret that. Sean, maybe you have a better way to interpret it, but I thought that was really unstable, and I question the scientific acceptability of the composite.

DR. O'BRIEN: Well, I think if you take this in context to other measures that

are out there and being used, it is actually a measure that has a fair amount of precision. We compared the outlier status when we are classifying hospitals based on the same rigorous Beysian probability, just using mortality, and there is basically no discrimination.

Mortality is a measure that is well accepted, widely used, and it basically did not discriminate nearly to the extent that the composite does. Now part of the reason why the second year, when you are comparing what you see in one year and the second year, is that in the second year you also -- based on this 99 percent probability.

Now a few of us said the second year, do we believe that hospital was above average or not. You know, based on a 50 percent criterion, you are really going to see a situation where hospitals where you are 99 percent certain in the first year that they are above average, and now we think they are

below average.

Basically, our best bet, you know, is that these hospitals remained above average. We weren't 99 percent certain. We went from being 99 percent certain to slightly below 99 percent certain, but we still made it then 95 percent certain, 90 percent certain. So in that respect, I think it was still in okay shape.

CO-CHAIR FLEISHER: We need to move on. Is there any last new comment?

MEMBER HOPKINS: I just want to make a final comment on the issue that -- I think it was Joyce brought up. If you are going to tie the star rating system as proposed and it is built on 99 percent confidence, I can't in good conscious vote for that as --

CO-CHAIR FLEISHER: You will have a chance.

MEMBER HOPKINS: Okay. I'm just making a statement for others to think about.

I think the consumer should decide on what level of certainty they want, and certainly not that level. It is out of step with every other measure I can think of.

DR. O'BRIEN: We have -- Dave Shahian can speak to this, but the Society of Thoracic Surgeons got a lot of input from users of the measure and addressed the topic of using basically custom developed probabilities other than 99 percent for some applications, and you don't need to be 99 percent certain.

There's discussions of multiple payers about whether this particular system was working for them, and the feedback was universally positive. So that level of probability was working for the users to the extent that they were able to tell.

The users are participants who are receiving these as internal feedback report and they are third party payers who are gathering, basically, their data, voluntarily

reported them to the Society for Thoracic Surgeons.

CO-CHAIR FLEISHER: Other issues unrelated to the star rating? B.J.?

MEMBER TURNER: Just an observation, that we are giving these guys a pass, which I think is fine on process measures, like prescribing a certain drug. I'm just saying that, if that is the way we get them in, them I am all for it. So that is one point.

I would appreciate when we are going to this more discrimination in telling us about the exclusions, because it just says whatever, contraindications to XY, and I am trying to make sure that this is really covering the bases. So in the future I would like to know more about what you call an exclusion.

CO-CHAIR FLEISHER: Dianne.

MEMBER JEWELL: So, actually, I was going to point to this measure as a nice

example of how measure developers can specify a composite that is outcome and process based, and that is part of the issue, is we really need the measure developers to be able to well specify what their intent is, and then we also have to be consistent in our scope.

I think those are -- That is what we have done here. We are not passing anybody. We have been clear about our scope, and we have been clear about integrity of process, and we are clear that the measure developers have given us something that reflects both of those things appropriately.

CO-CHAIR FLEISHER: Thank you. So time to vote.

DR. WINKLER: All right. Four criteria and then the recommendation. So for composite measure for CABG, procedure.

CO-CHAIR FLEISHER: This is criteria voting.

DR. WINKLER: So, importance to measure and report: Yes?

CO-CHAIR FLEISHER: Uniform.

DR. WINKLER: Okay. So it is 22.

All right. Scientific Acceptability of the measure properties: Completely meets criteria for scientific acceptability?

CO-CHAIR FLEISHER: Well, I think you need to vote -- I think you could place it in either. If you do not believe it is scientifically valid, then it should be voted partial or none.

DR. O'BRIEN: Currently, it is reported as the actual number along with the confidence interval. In addition to that, there is a star.

MEMBER HOPKINS: That is important

CO-CHAIR FLEISHER: Do you want to clarify that?

MEMBER HOPKINS: That is the number in the confidence interval and the star.

MEMBER AMARASINGHAM: But it is an

important point, in that I think it addresses the original concern.

CO-CHAIR FLEISHER: So we will -- according to the group up here, we will say that the star rating is part of usability, and therefore, we do not include that in the scientific acceptability. Is that okay, to make a comment as part of the voting?

MEMBER HOPKINS: What we are voting on has only the star rating?

CO-CHAIR FLEISHER: No, no, no. No, what I am saying is that we will, as part of the definition of our voting, say that we consider the star rating part of usability and not scientific acceptability. We will put that condition on our -- So if you vote, you will know that it has this comment that goes forward up the chain.

DR. BURSTIN: It is specifically under the usability section for this discrimination.

CO-CHAIR FLEISHER: Go ahead.

DR. WINKLER: Okay. For scientific acceptability, completely meets the criteria? I get 15.

Partially meets criteria? Five.

Minimally? -- Oh, you have six? Okay, I missed one. All right.

Usability: All right, completely meets? I get a zero.

Partially? Fifteen, okay.

Minimally? Three.

Not at all? One -- I am still missing one. It was 17 on the partial? Okay.

Now feasibility: Completely meets criteria? Eight.

CO-CHAIR DUBOW: Withdraw one?

DR. WINKLER: Whose?

CO-CHAIR DUBOW: Mine.

DR. WINKLER: I didn't count you yet. Okay, Partially? Twelve.

Minimally? That should be one. Okay.

CO-CHAIR FLEISHER: And you must -

- You must be part of the registry to get this data as opposed to other measures.

DR. O'BRIEN: Dave, feel free to answer the registry conclusive -- we think 90 percent of the hospitals in the U.S. have a marginal cost, separate from the fact we already are collecting this data, getting all kinds of reports. So the additional cost of doing this composite is -- you know, there is no additional cost.

MEMBER HOPKINS: That is fine if you want to report at the hospital level. If you want to report at the surgeon level, what then?

DR. SHAHIAN: We don't calculate this at the surgeon level.

CO-CHAIR FLEISHER: But this could not be calculated independently of joining the registry?

DR. SHAHIAN: That is correct.

MEMBER AMARASINGHAM: When you use the 90 percent, though, you know, a lot of

hospitals don't perform bypass surgery. So does that 90 percent include 90 percent of those that perform bypass surgery or 90 percent of U.S. hospitals?

DR. SHAHIAN: Ninety percent of the hospitals that -- Yes, about 90 percent of the hospitals that perform cardiac surgery are in the registry.

MEMBER YAWN: Yes. It is nowhere near 90 percent of hospitals that perform cardiac surgery.

MEMBER GERBIG: Will it continue to be correct that surgeons can join. Surgeons and hospitals can join, but hospitals on their own cannot submit?

DR. SHAHIAN: No. A hospital can be an STS participant. A surgical group can be an STS participant, and in rare cases an individual surgeon, and that is very rare.

Most commonly, it is a hospital submitting its entire cardiac surgery results, but occasionally it is a large surgical group.

CO-CHAIR FLEISHER: Okay. Now for the vote. Any conditions that anyone wants to put on the vote? Do you want to propose your condition? Yes or no?

MEMBER KEALEY: I am still confused by the star system. I don't quite get it, why it has to be there.

CO-CHAIR FLEISHER: Well, that -- So you can propose that it is approved without the star system? Are you proposing that?

MEMBER KEALEY: I am hoping that maybe they can explain it one more time so I can understand what exactly this is. I thought it was just the stars, but then it sounded like the numbers are being reported, and so it is an either/or thing.

DR. SHAHIAN: The problem is that the numbers for many people would be very difficult to interpret, and as a way of making the measure more usable for the general public, we calculated what we believe to be a very responsible system for differentiating

truly superior programs, programs that are clearly having an issue, and the large percentage of programs, about 75 percent, that are statistically indistinguishable.

They may vary in certain characteristics in certain domains, but it is very hard to distinguish them statistically. Frankly, it is our belief, and we have had many external observers comment as well, that this is pretty consistent with how we view things clinically, that there are a few truly superior places, a few programs, 10-15 percent, that are operating suboptimally, and then the vast majority operating at a very high level and not able to be distinguished from one another statistically.

We think that works. It helps --

CO-CHAIR FLEISHER: I think we have the point.

MS. HAUGEN: Just a comment from a consumer standpoint. You know, one star, two stars, three stars, I think, would be

interpreted as good, better, best, not as --- I mean, one star here is below the mean, which means it is poorer than what -- at least the minimum we should be able to expect. That is how I interpret this.

I don't know how this -- whether this would be translated by the public on its own in accordance with what the data is really telling us, what the statistics are telling us. So that is, I guess, a concern I have.

As you look at this -- I mean, the layperson thinks, if they got one star, that is pretty good, maybe not as good as the third, but where I am maybe that is pretty good, and that is not what this is telling you.

DR. SHAHIAN: We never -- There is no STS publication of any kind that has ever said good, better, best or we would very definitely state that, that one star is an underperforming program.

MEMBER HAUGEN: Yes. I am just

saying -- I am taking it from saying the layperson that looks at this -- I am not saying that you have done that. But if you look at this in isolation, that is the way this type of mechanism -- and maybe there is a different way of visualization, because people will visualize this and interpret this in that way. That is my perspective.

CO-CHAIR FLEISHER: Okay, thank you. Any public comment before we vote? Okay. I am going to take Chair prerogative that we vote the measure as is, vote the measure with the condition that the star system not be part of the measure, the endorsed measure, and vote against the measure.

CO-CHAIR DUBOW: I just wonder whether we should ask the developer whether --

CO-CHAIR FLEISHER: We did, and he said maybe. He said he would have to go back.

MEMBER AMARASINGHAM: I think, if you vote for the condition, you should vote

recognizing that it could get rejected.

CO-CHAIR FLEISHER: If we vote for the --

MEMBER AMARASINGHAM: It means that progress on this measure might be --

CO-CHAIR FLEISHER: Do we want to vote -- Are you proposing we vote that they respond to this and we re-vote, if they say no?

CO-CHAIR DUBOW: No. I think what we could do is -- I think we could express -- I think we could have an up or down vote on the measure as it is currently presented, and express some concern about the fact that this reporting mechanism is tethered to the measure itself.

My guess is that there is some sentiment that the reporting mechanism is less than ideal, doesn't sound as though it has been tested among all users, just professional users, but that the measure itself clinically has a great deal of validity and importance.

So I would prefer to express some concern rather than to see the measure go down.

CO-CHAIR FLEISHER: Okay. So does anybody -- Sean, you want to make a comment?

DR. O'BRIEN: I think a can of worms was opened by accident that didn't need to be opened, and worms could be put back in.

In the submission form that I wrote, it says the STS CABG composite score. Now the STS CABG composite score is the backbone for the star rating system.

It is used by the Society of Thoracic Surgeons, but I think everyone would be very happy if there was a vote made on the STS CABG composite score as a submitted measure, and then no one needs to discuss the star. You are happy to have implementation issues be a separate implementation issue. Let it be a separate implementation issue, and you don't need to specify we don't like the star rating system, we do like it, and it is

a separate issue.

CO-CHAIR FLEISHER: Yes. Okay. Yes, David?

DR. SHAHIAN: I would just want to know what the implications of that were, because we have worked very hard to create the star system, to have it in operation for several years, and if this vote means that the star system has no NQF endorsement, then I think we would want to come back and talk to you about that.

DR. O'BRIEN: I don't think anyone is going to be -- It is going to be that the STS CABG composite score is endorsed, and that is going to be what is desirable.

CO-CHAIR FLEISHER: Is that --

DR. O'BRIEN: I suspect that you wouldn't be in a law suit against STS if six months from now that the score itself was not publicly reported. This score --

DR. SHAHIAN: You know, it makes it really awkward for the STS, which reports

this and is going to report it publicly using the star rating -- It makes it very awkward for us to try to explain, well, the STS composite score is endorsed, but the star rating system, which is what you are seeing, isn't endorsed. That is very awkward.

CO-CHAIR FLEISHER: Okay. So what I am hearing is a proposal to have two separate votes. One is to endorse the measure as written and, two, to separately vote on the star system; and the measure developer can deal with -- There is still public comment. there is still responses, and there is still CSAC. Is that what you are proposing, Joyce?

CO-CHAIR DUBOW: I think what Sean is proposing makes sense. If we endorse a measure with the STS score, that seems to be the guts of the methodology. The reporting is -- we have not taken -- We don't -- So I think that whatever STS decides to do on reporting -- So just severing the two seems to be acceptable to Sean. I think he proposed it.

CO-CHAIR FLEISHER: Then do we vote on --

CO-CHAIR DUBOW: No, no.

CO-CHAIR FLEISHER: Well, the question is would we then not vote or vote, vote on the reporting mechanism, since we have suggested that we don't endorse reporting mechanisms?

MEMBER YAWN: Can't we just make a comment? I mean, we vote it up or down as -- not the reporting, just the measure, and then re have every right to make a comment saying we don't like the star system, we don't believe it is as transparent as they think it is. End of story.

CO-CHAIR FLEISHER: And we will take a simple vote on that to give the strength of that.

MEMBER HOPKINS: As a separate question we haven't dealt with.

CO-CHAIR FLEISHER: Yes.

MEMBER HOPKINS: So, Reva, the

measure we re voting on is just a simple measure. Right? The STS composite? Are we voting on it exclusively as a measure of hospital performance or more generally that could be applied to physicians, if they were willing? That is my question.

DR. WINKLER: This is submitted at the practice -- of either the group practice or hospital level. It was not submitted as an individual surgeon measure, and that is what we are evaluating it for.

MEMBER HOPKINS: I'm just trying to understand the implications. No one could take it -- Even STS couldn't take it and apply it to an individual surgeon, if they wanted to and say it was an endorsed measure.

CO-CHAIR FLEISHER: The latter half is the important condition. They could do it, but it wouldn't be endorsed.

MEMBER HOPKINS: I understand the restriction. We haven't heard it. I understand sample size issues, but --

CO-CHAIR DUBOW: The measure developer didn't propose it.

MEMBER HERMAN: And most places, when they look at it, do do it by the individual surgeon level, because if you have an issue with your measure, you have to have a place where you can go.

CO-CHAIR FLEISHER: And actually, cardiac surgery is a group sport, because it is a team, the ICU, the ward care, the nursing care, the perfusionist. So let's vote.

So we are going to vote on endorsing the measure at the practice level or hospital, yes or no, simple vote. Then you are separately going to vote as just a simple recommendation, and the strength of that recommendation with regard to the star rating.

So how many vote yes for this measure?

DR. WINKLER: It is everybody.

CO-CHAIR FLEISHER: Okay. Secondly, how many people vote that the star -

- the implementation using the star system -- how many people vote that they should reconsider that? I don't know if you want to propose --

MEMBER AMARASINGHAM: A clarification. It is going to be reporting with the star system, the point estimate and the confidence interval? That what my understanding was. Right, Sean?

CO-CHAIR FLEISHER: The reporting actually is the point estimate --

CO-CHAIR FLEISHER: It would be everything in the form right now.

DR. O'BRIEN: Dave Shahian should speak up, but I think that the idea -- It sounds like STS is interested in receiving endorsement for the whole package -- sorry, for the star rating system. So it sounds like there is -- can the endorsement apply to the star rating system as well?

DR. SHAHIAN: Well, we would report the point estimate and confidence

intervals. We will continue to use the star rating, and you know, what you decide to do in terms of a recommendation is fine, and we will certainly take a look at it, but we have so much experience with this now that I think we will likely continue it.

CO-CHAIR FLEISHER: So I am not even going to take a vote. I am going to propose, unless anybody disagrees, that there will be a comment that significant concern was expressed by the committee of the utilization of the star system for implementation, and just leave it at that.

That doesn't say significant concern was expressed. Anybody disagree with that statement? That doesn't mean some people endorsed it. Okay.

Let's go to the final measure -- What?

DR. WINKLER: Well, actually, we have three more.

CO-CHAIR FLEISHER: No, no, no,

but I mean the final cardiovascular measure on post-operative stroke or death after carotids.

DR. WINKLER: Is somebody from the Society of Vascular Surgeons with us?

SVS REPRESENTATIVE: Yes, Josh --

DR. WINKLER: Thank you. Okay. So this measure, which is 0T1-011-09 -- This is the percentage of patients without carotid, neurological, retinal symptoms as a baseline, within 12 months immediately preceding their carotid endarterectomy who then experience stroke or death after undergoing the procedure.

CO-CHAIR FLEISHER: The year is the pre-op period.

DR. WINKLER: Right. That establishes the baseline.

CO-CHAIR FLEISHER: I thought it was 30 days.

DR. WINKLER: I wanted to say it is, too. I have to go look.

CO-CHAIR FLEISHER: Can the

developer comment? What is the post-operative surveillance for stroke?

SVS REPRESENTATIVE: It was intended to be in hospital.

CO-CHAIR FLEISHER: In hospital? Thank you.

DR. WINKLER: During hospitalization.

CO-CHAIR FLEISHER: In fact, that was the comments from the TAP.

MEMBER YAWN: Could we ask how long is an average stay in the hospital.

CO-CHAIR FLEISHER: Two days. Two days, I'm sure. So, Barbara, we need the microphone, but the question was how long do they stay in the hospital, and is that an appropriate measure?

If I remember the comments from the TAP, there were some comments. So, Ted, are you still there?

MEMBER GIBBONS: yes, I am.

CO-CHAIR FLEISHER: Can you

comment?

MEMBER GIBBONS: Yes. I think we were thinking that the likely time of observation should be longer than the in-hospital stay, because the time may need to be extended. The recognition of the stroke is of paramount importance, because these are individuals who enter surgery asymptomatic.

CO-CHAIR FLEISHER: Okay. Comments? Barbara?

MEMBER YAWN: Do we have data on what is the rate of stroke within 24 or 48 hours and within 30 and 60 days?

DR. HERMAN: Yes, we do.

MEMBER YAWN: Okay.

CO-CHAIR FLEISHER: Was there concern from the TAP that this was acceptable with this short a time frame?

MEMBER GIBBONS: I think, in terms of the ability to define it based on stent data, that that would be --

If I could comment, one of the

major concerns in the TAP was the differences in the evolution of practice patterns for carotid disease being treated with stenting versus surgical endarterectomy, and that the decision to do so was increasingly made, and justifiably so, by the surgeon, who may not only be involved with direct surgical endarterectomy but with the placement of the stent in collaboration with interventional radiology or interventional cardiology, and that the measure itself may need to be revisited as the practice pattern changes fairly quickly over the next few years.

CO-CHAIR FLEISHER: Other comments?

MEMBER PINDOLIA: I'm sorry. I don't see that figure in there. I still don't see 24 hours post, 48 hour or 30 days post. It just has in the hospital is 1.3 percent, and then 1.7 percent, but it doesn't have the extrapolation of how far out DC.

CO-CHAIR FLEISHER: Any other

comments, because we have a public comment. No? Please.

DR. JEWELL: -- and that one of the issues is who is making the assessment and the determination that there has been a stroke, that sometimes the reliability of the surgeon making the diagnosis is not the same, and the studies have shown that you get different rates, depending on who you have making the diagnosis and whether they are using a standardized scale or whether it is a neurologist or an internist making the diagnosis, and that this as an outcome measure from a consumer perspective makes a difference who is making the diagnosis to report the outcome.

CO-CHAIR DUBOW: Do we have a reaction from the developer? Could the developer respond to that, please?

SVS REPRESENTATIVE: Yes, thank you. First, let me just respond to the question about the percentage of strokes that

occur within 30 days, that occur in the hospital.

In our registry it is about 94 percent. So that it does -- In-hospital data really captures quite well the stroke rate, post-operative stroke rate, and our concern was that the data would be unreliable at the 30-day time point.

In terms of the question about who is making the assessment, there have been studies that show that, if a patient is in a randomized trial or a neurologist sees the patients that the rate of detection of small strokes is higher. However, hospitals are incented to record post-operative stroke, because it increases the complexity of the patient and, therefore, the billing.

Other measures that currently are in existence at CMS, such as their measure 166 which is stroke after cardiac surgery or after coronary bypass -- there is no specification made in that measure at all about how the

stroke is measured. In fact, it isn't even well defined. So at least by current measures, we think we are meeting the same standard.

MEMBER AMARASINGHAM: Let me just make sure I understand it. Are you suggesting that there is an incentive to up-code?

SVS REPRESENTATIVE: Yes.

MEMBER AMARASINGHAM: So that is a legal --

CO-CHAIR DUBOW: You are accurately coding.

SVS REPRESENTATIVE: Accurately is what I mean.

MEMBER NEWCOMER: His point was that hospitals had a legitimate reason to search for legitimate strokes because they would be up-codes. So there should be a higher capture rate, is all, not that they are defrauding.

CO-CHAIR DUBOW: No, I think it responds to the question, that there is an

incentive to capture them properly.

MEMBER FILLIPO: But the hospital can't code it until the physicians document it, that it is a stroke. I mean, the coder can't go through and say it is a stroke. I mean, the physician has to document that it is a stroke.

MEMBER DELLINGER: But it is a Medicare rule, but the hospital can certainly prompt the physician to make sure the diagnosis is made. They can query the physician. That is certainly done in my hospital.

MEMBER FILLIPO: Absolutely.

DR. JEWELL: Except now you have the opposite incentive. If it is going to be reported as an outcome and be a negative, then it goes against how much difference are you going to get for the DRG.

CO-CHAIR DUBOW: Well, look, I thought your question was whether these strokes were going to be properly reported.

The response was that indeed there is every incentive for them to be reported properly, because there is an incentive to do it. So that seems to answer the question that you posed.

DR. JEWELL: The TAP brought it up, that there would be the counter-incentive as well.

CO-CHAIR DUBOW: Okay.

MEMBER KEALEY: I was just going to say, this creates the other side of the coin, and as frequently we see, there is competing interests with payment and accurate diagnosis. I see this in health grades all the time with how thorough these document post-op issues, because they will be seen as complications and make you look bad, but they might improve your payment. What do you do?

MEMBER NEWCOMER: Your hospital administrator has an easy answer for that.

I want to comment about the denominator. It simply says carotid

endarterectomy, but there is very high interoperative variation in the indication for this test -- or for this procedure, and we don't adjust for that anywhere that I can see.

I am a little worried about that denominator being so vague that we won't get a decent result. So I would argue against the scientific validity of that measure.

MEMBER DELLINGER: Well, it is not every carotid endarterectomy. It is carotid endarterectomy in a patient who has had no logic symptoms for 12 months before the operation.

MEMBER NEWCOMER: You can't tell that from the coding. Well, I guess you can with the G code. Okay.

CO-CHAIR FLEISHER: Barbara?

MEMBER YAWN: I'm fascinated by the fact that people believe, with apparent great accuracy about the ability to say this patient has been totally asymptomatic, but they don't think that they can reliably

identify strokes in 30 days.

I'm sorry. That doesn't wash with me at all, and I would be very interested in 60 and 90 days out and what those data are, too, because if these were truly asymptomatic patients, you do something and now they have a stroke in the next year, they should be at quite low risk for that.

CO-CHAIR DUBOW: But the measure we have before us is short frame measure --

MEMBER YAWN: I understand.

CO-CHAIR DUBOW: -- and we just heard that that accounts for 90 percent of the strokes under these circumstances. So --

MEMBER YAWN: Up to 30 days, yes, we heard that.

CO-CHAIR DUBOW: Right. So that is the measure we have before us.

MEMBER YAWN: I understand. I am just suggesting there might be reasons not to vote in favor of it.

MEMBER KEALEY: Is that a

different -- The 94 percent is kind of a national average. At a place where they don't do them so well, suddenly that becomes 80 percent. So is this a differentiated that we are automatically getting rid of?

CO-CHAIR FLEISHER: Okay. Lee, last comment.

MEMBER NEWCOMER: Sorry, but could the developer tell me on how they would choose these carotid endarterectomies? What is the criteria for this patient being an operative candidate? It is not there.

SVS REPRESENTATIVE: Really, that underscores why this measure is so important, because neurologists and surgeons together make decisions to recommend this surgery or perhaps in the future stenting when that is approved for asymptomatic patients, based on the severity of this analysis and all the other factors that would influence morbidity and, therefore, the outcome has to be really good.

That is why we think it is so important to keep track of this, even at the in-hospital level, which would provide the consumer, I think, with great data that is just simply not available now.

MEMBER NEWCOMER: I am going to argue this surgery is a morbidity, and you could definitely have your post-op morbidity decline by operating nothing but healthy carotids -- I mean, to be extreme. I just can't find validity.

CO-CHAIR FLEISHER: It is time to vote on the criteria.

DR. HALL: Are you ready for public comment?

CO-CHAIR FLEISHER: We opened it up. Is there other public comments?

DR. HALL: This is Bruce Hall. I just had a question on the measure specification. Again, it is a question I have asked repeatedly during the day.

I do not see data on reliability.

The measure developers talk about information like this. We do see rates from about 4.1 percent, a 3.8 percent and so on. I know those numbers may differ in different situations.

A great proportion of the hospitals across the country are going to do a significant number of these cases, to the degree that they can be reliably assessed on this outcome that is probably going to be under 3 percent.

Furthermore, the measure is supposed to be eligible for provider physician level as well. How do we know what the reliability distinctions between physicians and how many physicians do enough of these cases to be reliably judged?

CO-CHAIR FLEISHER: Okay, thank you, Bruce. Reva, do you want to go over criteria?

DR. WINKLER: All right. For the criteria, for this measure of stroke after

carotid endarterectomy: Importance to measure and report: Yes? All those that say yes? All but Barbara and Lee. Okay.

So how many No? There they are.

So that's 20. All right.

MEMBER YAWN: That is as specified.

DR. WINKLER: Scientific acceptability of this measure as specified: Completely meet the criteria? Two, okay.

Partially meet the criteria? Nine.

Minimally meet the criteria? Nine.

CO-CHAIR FLEISHER: Doesn't meet the criteria, two.

DR. WINKLER: Definitely on the low end.

Usability: Completely meet the criteria? I see zero. I'm sorry, David. So one.

Partially meet the criteria of

usability? Thirteen.

Minimally meet the criteria of usability? Six. That's 20.

Not at all? Two, okay.

Feasibility: completely meets the criteria? Three.

Partially meets the criteria? Twelve.

Minimally? Five.

Not at all? Okay. All right.

CO-CHAIR FLEISHER: Call for a vote. Any conditions? No. Those in favor of the measure?

DR. WINKLER: Five.

CO-CHAIR FLEISHER: Those opposed to endorsing the measure?

DR. WINKLER: Seventeen. That's everybody.

CO-CHAIR FLEISHER: Thank you. Okay, last two measures as the day continues. Bruce, you are still there, obviously. We are on to the two ACS measures.

The risk adjusted colorectal surgery outcomes measure. Was that one of the other times?

DR. WINKLER: Yes, it actually was.

CO-CHAIR FLEISHER: It was done in the GI TAP?

DR. WINKLER: Yes.

CO-CHAIR FLEISHER: David, did you -- Was David the Chair of that?

DR. WINKLER: Yes.

CO-CHAIR FLEISHER: David, do you have any comments on that measure from the TAP's perspective?

MEMBER JOHNSON: The measure was viewed favorably.

CO-CHAIR FLEISHER: Not on mike.

MEMBER JOHNSON: The measure was overall viewed favorably. It was defined need, and the only question was the participation in the Ethnoscript database, and that there was a mechanism to account for

participants that maybe weren't participating in the Medscript database. So that the developer addressed that and provided a pathway for that. So overall, the impression was favorable.

CO-CHAIR FLEISHER: So just to recognize that, unlike the STS measure, the ACS measures -- the hospital does not have to be a participant in the registry.

Any questions? Patch?

MEMBER DELLINGER: Yes. The STS criteria are very rigorous and not that easy to do. So I am quite -- sorry, NSQIP. NSQIP criteria are very strict. I am quite curious as to how a non-NSQIP hospital could get measured by this NSQIP criteria.

CO-CHAIR FLEISHER: Bruce, I know you have answered this before. Would you -- but not for the whole Steering Committee. Do you want to address that?

DR. HALL: Sure. Thank you. I have been on the call all day, and I have

appreciated hearing all of the discussion up to this point.

The measure before you is a very parsimonious measure. It is specified with a small number of data points, and it is specified with a subset of colorectal -- and also a small number of outcome endpoints. So any implementation of the measure would be accompanied by education about how each of those risk factors or outcomes is defined and applied, but the measure before you is a very parsimonious model, and we have given very specific estimates of what we think the reporting value would be.

MEMBER JOHNSON: The other point that was recognized is that, although the NSQIP participants weren't necessarily uniform 100 percent, the people that were doing the colorectal surgery represented 85 percent of the eligible surgeries already. So the majority of surgeons were already participating.

CO-CHAIR FLEISHER: Other comments? David?

MEMBER HOPKINS: The measure as stated is for what level of measurement? Hospital, group, physician?

DR. HALL: Institutional.

MEMBER HOPKINS: Hospital only.

CO-CHAIR DUBOW: Page one, the summary provides the level of analysis, thanks to Heidi.

CO-CHAIR FLEISHER: Other comments? Public comment? Reva?

DR. WINKLER: Okay. So importance to measure and report on a measure of colorectal surgery; so importance to measure and report, Yes? Is that everybody? Is there anybody voting no? Okay, good. All right.

Scientific acceptability of the measure properties for this measure as specified: Completely meets criteria? Eleven.

Partially meets criteria? Eleven.

That is everybody.

Usability: Completely meets criteria? Criteria for usability. Eleven, completely.

Partially? Eleven. So that is everybody. Okay.

The last one is feasibility: Completely meets feasibility criteria? Seven.

Partially? Thirteen.

Minimally? Two. Okay.

CO-CHAIR FLEISHER: Conditions before we vote? None. Hearing none, all those in favor of endorsing this measure? It is unanimous.

DR. WINKLER: Okay, great.

CO-CHAIR FLEISHER: Thank you. We have one measure left. This is a similar mini-NSQIP measure. It is risk adjusted care mix adjusted elderly outcomes measure developed by the American College of Surgery. It was discussed extensively at the TAP, and essentially, this is a mixed group of surgical

procedures similar to the previous measure. It was actually -- Bruce had explained the way that non-NSQIP hospitals could actually calculate it, because I believe there is only two criteria for risk adjustment. So it is relatively simple. Did I get that right, Bruce?

DR. HALL: It is a very small set. I think there are -- In total, if you include demographics and what-not, you are talking about half a dozen factors.

CO-CHAIR FLEISHER: Right.

CO-CHAIR DUBOW: Could you tell me if this includes hip fractures?

DR. HALL: Yes, that was a particular question during the TAP. So at first NSQIP approaches, multi-trauma and severe trauma patients are not eligible to take part in NSQIP, but isolated trauma such as fall from standing or slip from standing that might be associated with a hip fracture, which is relative to this population, would be

included.

So the way to think about it in the shorthand, it is just that, if you fall from standing and fracture your hip, you would meet it. If you fall off the roof of your house and fracture a hip, then you are out.

CO-CHAIR DUBOW: Okay.

CO-CHAIR FLEISHER: The TAP actually felt, as far as gap in measures or future research, that a hip fracture measure independent of this measure should be developed, and they actually felt that quite strongly. So that -- Based on Joyce's question, if the Steering Committee also agrees with that, we will -- no, not as a condition -- a recommendation. Yes, Dianne?

MEMBER JEWELL: Actually, no. I would ask, given that the Bone and Joint TAP received no measures for consideration, and hip fracture is one of the specific conditions that is identified, I think it would be a helpful message to send that we really want a

specific hip fracture measure.

CO-CHAIR FLEISHER: So the question to Bruce -- I don't know if in NSQIP or if NSQIP could work with the orthopods to develop such a measure.

DR. HALL: Well, I will just reemphasize that this measure before you is based on the standard approach. So it does include, you know, at least the portion of the population that I described to you.

NSQIP does not approach severe trauma patients. There is a quality registry within the American College of Surgery, the trauma QIP, the TQIP which does more severe injury patients.

I don't know if I am speaking out of turn -- I don't think so, but certainly, the College would be happy to use the TQIP resources and to work with orthopedics in the future to that end, but I don't think that is feasible as a condition for the measure that is in front of you.

CO-CHAIR FLEISHER: No, no, no. I think what is being asked, Bruce, is that hip fracture, similar to colorectal -- we may want to pull out, and we are asking that you consider developing a hip fracture specific measure very similar to colorectal in the general measure, given its public health importance and Medicare issues. Barbara?

MEMBER YAWN: And I think that the idea of hip fractures only from standing or sitting is really a subset of the non-major trauma hip fractures. I mean, some of us were walking and went down one step and fractured something, not --

DR. HALL: Again, that would be included. If it is not a fall from height, that would an included case.

MEMBER YAWN: Okay. So your measure is broader than you suggested.

DR. HALL: I was trying to give a shorthand. So the fall from standing or walking, that is in; fall from height, off a

roof, that would be out.

CO-CHAIR FLEISHER: So the other Barbara.

MEMBER TURNER: I am interested in the breadth of surgical conditions that are being covered, and the fact that some institutions may have -- I don't know -- mostly neurological procedures, and other institutions might have thoracic CNS things.

CO-CHAIR FLEISHER: Right. So that actually was one of the major concerns of the TAP.

MEMBER TURNER: And?

CO-CHAIR FLEISHER: Bruce, comment on how you judge the severity of surgery, so to speak.

DR. HALL: Certainly. So first of all, there is a full CPT list of specifications for the measure, and that is submitted. I'm sure you are not looking at it at the moment. But the concern about standardizing across surgical procedures is a

very insightful one.

The approach we have taken to that is that we have dropped all of the CPT codes that are eligible for this measure into clinically related buckets based on their CPT coding. So in other words, two colorectal surgeries that are different versions or, for instance, would be in the same clinical bucket.

So we have identified roughly 135, 136 if you include the category of "other," clinical buckets that all the CPT codes fall into, and then each of those buckets is run as an initial regression against the outcome.

We take the results of that initial regression and generate a scale of risk score. That scale of risk score effectively gives you a measure of the endogenous risk of that procedure.

So a urologic procedure all by itself, all other things equal, has a very different risk than a hip fracture or a

colorectal procedure. But each procedure bucket is given its own scale of risk score, and those scores are then used in re-regression in conjunction with the other risk adjustment variables and in conjunction with the relative value units of scientific code.

So what that gives us is an ability to standardize across procedures. So if one institution did a procedure that had endogenous the lowest and another institution has high risk procedures by definition, then we control for that mix.

That is why the measure is referred to not just as patient risk adjusted but also case mix adjusted for the institution.

CO-CHAIR FLEISHER: Sean -- Well, you reviewed this, but let B.J. finish.

MEMBER TURNER: Well, just to respond. The same issues come up with how many populations you have looked at, your case mix stratification approach. Is it just one -

- same idea as we went through with the severity of illness measure issues.

DR. HALL: So the measure in front of you has been developed using historical NSQIP data, even though participation in NSQIP is not required to fulfill the measure. So all of the coefficients and the reliability assessments and so on are all developed from historical NSQIP data.

CO-CHAIR FLEISHER: Okay, Sean, do you have any comments? You did review this?

DR. O'BRIEN: Yes, I did review it. In terms of my comments, I didn't see much in the way of assessing calibration of the model. They assessed calibration with a Hosmer-Lemeshow statistic, but they also commented that with such a large sample size, a significant Hosmer-Lemeshow test p value, which would ordinarily indicate lack of fit, is not meaningful, because no model is actually literally a perfect fit. So often Hosmer-Lemeshow is just a measure of how large

your sample is. But even when you have -- when the Hosmer-Lemeshow is not meaningful, there's other ways to assess calibration graphically.

Typically, very often you will report comparisons of the observed to expected outcomes within subgroups or by deciles of predicted risk, and I was curious to know if the developers had done those types of analyses.

Because the populations for these two measures are relatively broad, assessing fit, not just globally, that also important subgroups would be warranted, and also this will be in the DRG group, since the DRG basically entered into the model is a two-step estimation procedure where they estimate -- you know, they basically get an estimate for each DRG, but that goes into the model in the second stages, kind of a linear term where you are basically assuming that, if you have a -- You have a continuous variable, and basically

you are making linearity assumptions. So it would be worth assessing that.

CO-CHAIR FLEISHER: Bruce?

DR. HALL: thank you. Thank you, Sean. Your remarks are always very well thought out. So in fact, we have given up -- Internally, we have given up using the HL fit statistics for exactly the reasons that we saw and the reasons that you reiterated.

So in fact, we do for all of our models perform fit curves where we describe -- we actually evaluate fit within that graphically, and that helps us to know where we are on the fit.

We provided the HL fit, the 50-50 codes it seems that many times people look for. We have publications that have already been published, as well as an additional publication under review that specifically evaluates the additional value of using scale of risk from CPT codes, and I appreciate you said DRG codes, but we are referring to CPT

codes.

We find that in all cases in applications like this, the application of scale of risk scores derived from CPT codes dramatically improves the fit.

So we did not submit any graphical pictures of the fit codes for this model. We would be happy to do that. We did not know that would be the question, but we do do that. that is our normal approach nowadays.

CO-CHAIR FLEISHER: Any comment?

DR. O'BRIEN: No. An unrelated comment, and I may have missed some of the discussion. But the models are highly parsimonious, and I understand the need to reduce the reporting burden.

This measure is going to be implemented beyond NSQIP, but the criterion for assessing the performance of the model has to do with reducing or eliminating or reducing to the extent possible bias due to differential case mix across the sites.

So you might think about treating a larger model as a gold standard and using that model to assess the extent to which case mix does vary, and the extent to which different case mix would produce bias, and then to what extent does a three-variable model or a six-variable model succeed at removing the bias that you think may be there based on what you have estimated from your larger model, just any way to validate the use of a three-variable or six-variable model.

DR. HALL: Thank you, Sean. Actually, in this month's American College of Surgeons we have a publication that evaluates the five-variable models with a specific procedural grouping, and as you have said, this is a topic that we continue to investigate. So we certainly agree with your remarks that these are important aspects to continue to investigate.

CO-CHAIR FLEISHER: Can you send that to Reva, if you get a chance, a copy of

that?

DR. HALL: Sure thing.

CO-CHAIR FLEISHER: David?

MEMBER HOPKINS: Just a question. So this measure is for over 65. Does NSQIP have in its armamentarium an under 65 measure that parallels this?

DR. HALL: We have not submitted an under 65 measure that is otherwise specified the same. We developed this measure in conjunction with consultation with CMS leadership, because they were particularly interested in the unique burdens and risks that the over 65 population has. So that is why we have taken this approach.

I don't know that there is any reason to preclude making a similar model for an under 65 population, but we have made the argument that the over 65 population carries uniquely increased risk and uniquely increased risk of morbidity after these surgeries.

CO-CHAIR FLEISHER: So, David, I

assume we -- One potential request is that this be looked at in the future.

MEMBER HOPKINS: Sure.

CO-CHAIR FLEISHER: Patch?

MEMBER DELLINGER: This is -- I think this is a great model, but I really fail to see how this could be applied to a non-NSQIP hospital.

DR. HALL: Dr. Dellinger, the CPT codes that would be eligible for data collection are specified. The number of cases that would need to be evaluated over the course of the year, I believe, is approximately 180 cases, and so a hospital could sample 15 cases a month.

The risk adjustment model gave a very parsimonious, fewer than half a dozen, risk adjustment variables. So any implementation would again be accompanied by education about how those fields are defined, and then whether the hospital was participating in NSQIP or not, the data for

their cases would just be submitted to whatever organization was doing the implementation.

The models would be applied, and the evaluations would be returned to those institutions.

CO-CHAIR FLEISHER: Yes, B.J. We have seven minutes to dinner.

MEMBER TURNER: So I think -- I am curious. Do you report your results in age strata, because although you say the over 65, 66 is like a kid compared to 98. So I am wondering whether -- or how you report or address the issue of age in your model.

DR. HALL: We don't report it in a stratified manner, but it is included as a variable that we can use for risk adjustment. So even within -- Even given the fact that the entire population is over 65. So we don't report out 65 to 70, 75 to 80 and so on.

MEMBER TURNER: But you do adjust for how old they are?

DR. HALL: Within the group.

CO-CHAIR FLEISHER: Other comments. Amy?

MEMBER ROSEN: I have a comment on functional status. If a hospital doesn't have NSQIP, then they have to go in and look at the medical records for that? I just want to be clear on that in terms of usability or feasibility.

DR. HALL: This question also came up during the TAP, and actually, again because we do have 250 or 270 hospitals around the country that use that same variable, our experience within the program is variable is actually pretty easy found within nursing assessments of the patient, that this is an access that the nursing that the nursing assessment does usually cover. So that might not be immediately obvious, but that is our experience.

CO-CHAIR FLEISHER: Okay, public comment? Hearing none-- Yes, from CMS.

DR. HAN: I think I have more just a question about the implication of NQF endorsement here. Am I correct to assume that NQF endorsed measures, sort of like public available, is transparent, because from CMS point of view that we interested -- We are going to -- We are interested in using measures the NQF endorsed.

So for example, the STS measures that I have this question. What is the implication of you endorse the composite methodology, but you put it aside, the question of the star system?

So does that mean that CMS has -- there is a transparency of the methodology. We can use the measure, but CMS can decide how we are going to do the star system, how we are going to report? I would like to know what STS's concern is. That is a separate thing. Right?

CO-CHAIR FLEISHER: Yes. But ay issues with this measure, we can try to

address that separately, or do you want to wait for Helen?

DR. WINKLER: The question on that is, again I think what this committee was saying was the actual implementation where you determine stars and whatever is separate from the measure. Even though STS has wrapped them together, the recommendation of this committee is that they really should be looked at separately.

DR. HAN: Okay. So the same thing applies to the ACS measures, that I think it was David asking, you know, how he -- he wonders how this measure can be used outside of NSQIP.

So CMS is interested in the measure, but when we implement it, CMS will have some authority to require hospital to do that. So it is the methodology that we are interested in. It is not the implementation. So just want to remind you, when you vote on this measure.

CO-CHAIR FLEISHER: I am sure, though, during the comment period, the American Hospital Association, in particular -- I am sure we will hear from Nancy Foster or others about the burden.

DR. HAN: That is the -- Okay.

CO-CHAIR FLEISHER: So I think the TAP, from what I am hearing, the TAP -- Excuse me, the Steering Committee can comment that we are concerned about burden, and that public comment regarding potential burden of this measure would be important before endorsement at the next level. Is that a fair comment to add, unless anybody disagrees with that comment?

MEMBER YAWN: That goes to feasibility and usability.

CO-CHAIR FLEISHER: Separate from endorsement, because we can endorse a measure but say --

MEMBER YAWN: But say we think there is a feasibility issue.

CO-CHAIR FLEISHER: So let's vote on those components of feasibility. So if you are concerned about that, you should vote lower on feasibility and usability.

MEMBER AMARASINGHAM: One question for my own edification as an internist. What is the number of hospitals that are participating in this group?

CO-CHAIR FLEISHER: Small. You hear it. Two hundred thirty-eight, Bruce?

DR. HALL: Two hundred seventy.

CO-CHAIR FLEISHER: Two hundred seventy, and that actually is the -- out of 5,000. It is actually the major concern of why, at least I have heard, the College is looking at other ways to do this, because -- We are a NSQIP hospital, and the burden is large, and the costs are large, although interestingly, the STS costs are probably very similar.

So let's vote on -- It is not low. The cost is low, but the burden to collect.

So can we vote on the importance?

DR. WINKLER: All right. Let's vote on importance for this measure, which is a measure of outcomes of mixed surgeries for patients over 65. So importance, yes or no. Yes? Is anybody voting No? Okay, so that's good.

Scientific acceptability: Completely meets the criteria? No.

Partially? That's everybody.

CO-CHAIR FLEISHER: No.

DR. WINKLER: No? Anne? Amy? You are a minimum. Okay.

All right, usability: Completely? That is zero.

Partially? Twelve, I think.

Minimally? Ten. Okay.

Feasibility: Completely? Zero.

Partially? 10.

Minimally? 11. Okay.

So recommendation on the measure?

CO-CHAIR FLEISHER: Okay.

Recommendation of the measure. Now recognize we will have a strong comment that they should get public comment regarding the usability of this measure.

DR. WINKLER: It is going to happen, no matter what.

CO-CHAIR FLEISHER: But that we have concerns about implementation.

MEMBER TURNER: And a lot of statistical issues that we were discussing.

CO-CHAIR FLEISHER: Well, that is separate. So are there any conditions that anybody wants on the measure? Okay.

All those in favor of the measure? Did anybody vote No? Is there any No? No. Okay. So 22 to zero.

So, Bruce, thank you. The one comment, Bruce, is Dianne is going to be talking with you, because the TAP and the Steering Committee both feel there should be a separate hip fracture.

Any comments?

CO-CHAIR DUBOW: No, but we are going to reconvene tomorrow at 8:30. Your shuttle will pick you up at 8:10, and we are eating here, for those of you who decided to do that, at six.

MS. BOSSLEY: Right. And we do have at least one person going back to the hotel now and taking a cab. So if anyone wants to do that, we will get enough. I guess it is just Amy right now.

CO-CHAIR FLEISHER: So one quick question. Because many of us assume that it was a four o'clock end time, but we are further -- those of us who took the train. Do other people have like strict deadlines? So is the goal to -- if we finish at three, will people need to leave early? If we finish at 2:30, will they need to leave early?

(Whereupon, at 6:03 p.m. the steering committee was adjourned.)

H

I

Q

T